December 1968

USCEE Report 319

UNIVERSITY OF SOUTHERN CALIFORNIA

Estimating Linear Regression Parameters in the Presence of Non-White Gaussian Noise With Unknown Covariance Parameters

by

Gene Barry Goldstein

Portions of this work were supported by the Air Force Office of Scientific Research under Contract No. AF-AFOSR-68-1555.

[redacted]

ELECTRONIC SCIENCES LABORATORY

December 1968 USCEE Report 319

Estimating Linear Regression Parameters in the Presence of Non-White Gaussian Noise With Unknown Covariance Parameters

by

Gene Barry Goldstein

Electronics Sciences Laboratory
University of Southern California, University Park
Los Angeles, California 90007

Portions of this work were supported by the Air Force Office of Scientific Research under Contract No. AF-AFOSR-68-1555.

ACKNOWLEDGEMENT

For his interest and guidance throughout this research, I would like to express my gratitude to Professor P. Swerling. In addition, I would like to acknowledge the many discussions with Professor Swerling prior to the beginning of the actual lines of research described in this thesis. These discussions have greatly increased my knowledge of communication theory and estimation theory.

I would also like to acknowledge the suggestions of Professors T. S. Pitcher, I. S. Reed, T. Harris, R. A. Scholtz, and L. Welch during many informal discussions of this research. I am also grateful for the encouragement and suggestions from several fellow students, especially Mr. R. Sheinberg and Mr. D. Elliott.

This research was carried out while the author was a Member of the Technical Staff of the Hughes Aircraft Company - Aerospace Group, and a holder of a Howard Hughes Staff Doctoral Fellowship. The author would like to acknowledge this support as well as the encouragement and helpful suggestions of several staff members, notably Dr. H. Davis.

For the computational results the author is indebted to the staff of the System Simulation Laboratory of the University of Southern California. Portions of this work were supported by the Air Force Office of Scientific Research under Contract Number AF-AFOSR-68-1555.

Thanks are extended to Miss Jan Vaughn and Miss Lucy Faulkner for their patience and understanding during the difficult task of final typing.

In the performance of research of this type there are invariably periods of little progress and disappointment. I would like to acknowledge the contributions of my wife, Francine, without whose patience, understanding, and encouragement, this work could not have been done.

ABSTRACT

A number of problems that arise in radar and sonar applications can be regarded as parameter estimation problems, in which the desired signal, $f(t,\underline{\alpha})$, is imbedded in non-white, Gaussian noise. It is desired to estimate the unknown, nonrandom parameter vector, $\underline{\alpha}$, from observations (continuous or sampled) of the received noisy signal over a finite time interval $[0,T]$. Here $f(t,\underline{\alpha})$ is a known nonstochastic function, and we shall consider the case when $f(t,\underline{\alpha})$ is linear in $\underline{\alpha}$. In this case, $\underline{\alpha}$ is referred to as a linear regression vector.

We shall investigate the variance of the Least-Squares (LS) estimator and of the so-called Generalized-Least-Squares (GLS) estimator for $\underline{\alpha}$. Both are unbiased estimators for $\underline{\alpha}$.

When the noise covariance function is completely known one may construct a minimum variance unbiased estimator (MVUE) for $\underline{\alpha}$, and this estimator is a member of the class of GLS estimators.

Our interest is in the case when the noise covariance is not completely known, but may be regarded as a known function of a finite number of unknown, nonrandom parameters, $\underline{\beta}$.

It is shown that when $\underline{\beta}$ contains any covariance parameters other than the noise variance, there exists no MVUE for $\underline{\alpha}$.

However, we shall exhibit a class of problems for which the MVUE for $\underline{\alpha}$ has a variance which is orders of magnitude smaller than that of the LS estimator. In such a case it is of interest to find an estimator which makes use of whatever covariance information is available in an attempt to approach the performance of the MVUE.

It is shown that we can significantly improve upon the LS estimator by employing a bootstrapping procedure to estimate $\underline{\alpha}$. In some cases the bootstrapped estimate of $\underline{\alpha}$ can be shown to be unbiased. In any case, it is demonstrated via computer simulation that the bootstrapped estimate of $\underline{\alpha}$ is capable of reducing the variance of the LS estimate by orders of magnitude. In fact, the mean squared estimation error using the bootstrapped estimator for $\underline{\alpha}$ may be within a few percent of the variance of the MVUE, i.e., the variance the MVUE would have if $\underline{\beta}$ were known a-priori.

The bootstrapping procedure consists of using the LS estimate of $\underline{\alpha}$ to provide an initial estimate of the regression vector from which an initial estimate of the unknown covariance parameters is constructed.

Two procedures are outlined to accomplish the estimation of $\underline{\beta}$. The first approach is based upon an application of the theory of locally best unbiased estimation. The second approach is herein termed the "inverse-covariance-function" technique. Because of its simplicity, the latter approach is employed in the simulations.

Regardless of the manner in which the covariance parameters are estimated, these estimates are used to construct the GLS estimator for $\underline{\alpha}$. This is the first iteration of the bootstrapping procedure.

The GLS estimate of $\underline{\alpha}$ is then used to re-estimate the unknown covariance parameters, and then to re-estimate the regression parameters.

The process uses only the one available record of data, and may be repeated ad nauseam. However, dramatic results were obtained after only two iterations of the bootstrapping procedure.

CONTENTS

Chapter 1
INTRODUCTION

1.1 Problems and Objectives

A number of problems that arise in radar and sonar applications can be regarded as parameter estimation problems in which the data

$$r(t)=f(t,\underline{\alpha}) + e(t) \qquad (1\text{-}1)$$

are received (either continuously or sampled) over a finite time interval $[0,T]$. Here the desired signal, $f(t,\underline{\alpha})$ is a known function of the unknown n-dimensional parameter vector, $\underline{\alpha}$, and $e(t)$ is observation noise. We shall be interested in the case where $\underline{\alpha}$ is a nonrandom parameter vector and $e(t)$ is a sample function from a continuous-in-mean, zero mean, Gaussian random process. We shall restrict the discussion to linear parametric dependence. That is,

$$f(t,\underline{\alpha}) = \sum_{i=1}^{n} \alpha_i g_i(t) \qquad (1\text{-}2)$$

where $\{g_i(t) : i=1,\ldots,n\}$ are known nonstochastic time functions.

It is well known that with the above assumptions, the minimum variance unbiased estimator (MVUE) of the regression parameter vector, $\underline{\alpha}$, is a function of the covariance properties of the noise process.

In this dissertation, we consider the problem of estimating $\underline{\alpha}$ when the noise covariance (or equivalently the spectrum in the case of a stationary random process) is not completely known.

The method of approach is to treat the noise covariance function as a known function of a finite number of unknown, nonrandom parameters.

The main objectives of this work are:

1) to provide some insight into cases where dramatic improvement over the performance of the simple least-squares (LS) estimate†, which makes use of no knowledge of the noise covariance, is theoretically possible.
2) to propose a reasonable estimation procedure which is capable of offering significant improvement over the performance of the LS estimator.
3) to investigate the performance of this procedure analytically and experimentally via several examples which demonstrate its utility.

1.2 Summary of Previous Work

The subject of optimum* estimation of $f(t,\underline{\alpha})$ when the noise statistics are completely known has been investigated by many authors. Rao's book [1] is an excellent reference on the subject. Grenander

---

†See Section 2.2 for the definition of the LS estimate.

*In the sequel optimum always refers to minimum mean squared error.

and Rosenblatt [2] present the material more from the point of view required in the treatment of random processes than does Rao. Parzen [3, 4] uses the tools of Hilbert space to provide an approach which is applicable to both the continuous-time and sampled data cases. Cramer [5] and Wilks [6] are good references for an understanding of the basic mathematical statistics required in the study of parameter estimation problems. Swerling [7] derives useful expressions for evaluating the covariance matrix of vector parameter estimates for a wide class of problems. In another paper [8] he discusses various approaches to the parameter estimation problem, including the case in which the unknown parameters are regarded as having a known a-priori probability distribution.

On the subject of power spectrum and covariance function estimation the classical references are Blackman and Tukey [9] and Grenander and Rosenblatt [2]. A recent paper by Parzen [10] summarizes much of the present state of knowledge on this subject. All of these works are concerned with estimation of the entire structure of the spectrum or covariance function, however, and are not applicable to the case in which the covariance or spectrum is known but for a finite number of unknown parameters.

Two papers concerned with the parametric approach to power spectrum estimation are those by Levin [11] and Hofstetter [12]. Levin's

results are approximate results which hold under a wide variety of conditions. Unfortunately, however, the effects of the approximation have not been quantitatively evaluated. Furthermore, for the examples presented in this dissertation, where the optimum estimator is significantly better than the LS estimator, it can be shown that Levin's approximations are invalid. And while Hofstetter's approach is exact, he is unable to obtain any analytic results except for the case of estimating a spectral amplitude scale factor. The estimation of a spectral scale factor is not pertinent to the problem of estimating linear regression parameters, which we are concerned with here.

A recent paper by Rao [13] considers the linear regression parameter estimation problem when the noise covariance is unknown, and an estimate of it is incorporated into the regression parameter estimate. Here the noise covariance is regarded as completely unknown, and estimates are constructed for each element of the noise covariance matrix. Rao's assumptions require, however, that a number of independent realizations of the same random process be simultaneously obtained (e.g., via multiple sensors). (These separate realizations could be correlated if their correlation is known completely.) In fact, the number of realizations obtained must be at least as great as the number of time samples available in each realization. This follows from the conditions for the existence of the Wishart distribution which is central to his results. In his paper, Rao uses the

available independent realizations in order to construct the covariance estimates.

Up to this point, references have been cited which fall into one of three categories:

1) Signal parameter estimation with known noise covariance
2) Noise spectrum or covariance estimation (parametric and nonparametric)
3) Regression parameter estimation using an estimated (nonparametric) noise covariance matrix.

The parametric covariance approach to be adopted here really amounts to a particular kind of joint parameter estimation problem. As such, it is logical to ask if there exists some general theory which is capable of providing optimum estimators of the desired parameters.

The results of Barankin [14] provide the desired theory when one is willing to accept what are termed "locally best" unbiased estimates of the desired parameters.† In some cases only locally

---

†We are not really interested in locally best estimates of $\underline{\alpha}$. However, an application of Barankin's theory will reveal that under certain conditions (to be stated) no MVUE of $\underline{\alpha}$ exists. We will then abandon the requirement of bestness and hope to find an estimate for $\underline{\alpha}$ whose performance is close to that of the MVUE when the latter exists.

best unbiased estimates will exist. A necessary and sufficient condition for the existence of locally best and for uniformly best (i.e., MVUE) unbiased estimates was stated by Rao [1]. Barankin gives a different necessary and sufficient condition for the existence of a locally best unbiased estimate. Swerling [15] restates Barankin's results in a form more easily applied. Since neither of these papers has received much attention, and because of the author's own interest in this important subject, much of the development will be elaborated upon in this dissertation.

In addition to the above-mentioned references, there are a host of papers [16-33] in the fields of cybernetics and adaptive control which the author has found useful and stimulating.

1.3 Outline of Dissertation and Summary of New Results

Chapter II begins with an investigation of the simple LS estimation procedure and proceeds through a development of the generalized-least-squares (GLS) estimator*. The optimum estimator, when the noise covariance is known, is a member of the class of GLS estimators. For-

---

*The GLS estimator is defined in Section 2.3.

mulas are given for the covariance matrix of the parameter estimates. The results are given for the continuous-time and sampled data cases. These results are not new, but their development constitutes a convenient introduction to the subject and symbology which follow them. Using these tools, two examples are provided in which dramatic improvement over the performance of the LS estimator is theoretically possible. These examples are investigated in greater depth in the chapters which follow.

In Chapter III the Cramer-Rao bounds for joint unbiased estimates of the regression and covariance parameters are derived for the first example mentioned above. In this example the noise process is a stationary first-order autoregressive scheme. It is shown that the bound on the regression parameter estimate is not increased by the presence of the unknown noise covariance parameters. Since this bound cannot be achieved except under certain limiting conditions, an investigation into the Barankin bound, which is always achievable, is made. The Barankin bound analysis is quite general and is not restricted to the case of autoregressive noise. (An exposition of Barankin's theory is given in Appendix II.) It is shown that no uniformly best (i.e., MVUE) estimator for the regression parameter exists when the noise covariance (normalized) has unknown parameters. A technique employing barankin's theory for the estimation of pertinent covariance parameters is suggested.

Chapter IV is concerned with the synthesis and analysis of the desired estimation procedure. The procedure involves nonlinear operations on the data, making analytic results for bias and mean square error calculations difficult to obtain. Some analytic results regarding the bias of the estimator are presented. A detailed investigation into the performance of the procedure is made via a digital computer simulation of the two examples mentioned earlier. In the course of developing the proposed procedure, a comparison is made to the joint maximum likelihood estimator.

The main result is the demonstration (via simulation) that significant improvement over the LS estimator performance can be obtained in some cases by the use of the proposed procedure. The procedure actually amounts to an unsupervised learning or "bootstrapping" technique in which the LS estimate of the regression parameter is used to estimate the noise covariance parameters. The estimated covariance parameters are then used to revise the regression estimate making use of the GLS estimator. The process can be continued ad-nauseam. However, the examples studied indicate that after only two revisions of the LS regression estimate, the mean squared error may be nearly equal to the variance of the MVUE.

This result runs counter to a remark made by Eicker [34] that a useful estimate of the covariance matrix, or of the functions of it which are pertinent to estimating the regression parameters, cannot be constructed from a single finite sequence of observations.

Chapter 2.
LINEAR ESTIMATION

2.1 Introduction

By a linear estimator, $\hat{\underline{\alpha}}$, is meant a measurable function of the observables, $\underline{R}$, which is linear in the observables. The interest in linear estimators is due largely to the ease with which they can be synthesized and also to the availability of analytic tools to investigate their performance. Furthermore, for the linear regression model of interest here, the MVUE of $\underline{\alpha}$ is a linear estimator.

In this chapter the LS and GLS estimators of $\underline{\alpha}$, which are linear, unbiased estimators of $\underline{\alpha}$, will be investigated. The data will be assumed to consist of time samples, and vector-matrix notation will be employed for convenience. The generalizations required for treatment of the continuous-time problem will also be given.*

---

*In general the results for time sampled problems will depend upon how the time samples are distributed on the observation interval. This dependence will not be explicitly denoted in the sequel, however.

It is well to observe that if the time samples are constrained to be equispaced on an observation interval of fixed length, the estimation accuracy is not necessarily monotonically improved as the number of samples is increased. It can be shown, however, [35] that the optimum linear estimate is obtained for dense sampling in the interval. (Also, of course, a set of samples which contains another set will not lead to a worse estimate.)

2.2 Least-Squares Estimation

The problem of interest may now be stated as the estimation of $\underline{\alpha}$ when the data is of the form

$$\underline{R} = G\underline{\alpha} + \underline{e} \tag{2-1}$$

Here $\underline{R}$ is an Nx1 column vector consisting of the N time samples of data on the interval [0,T]. G is a known Nxn matrix whose elements are given by

$$G = [g_{\mu i}] \tag{2-2}$$

$$g_{\mu i} = g_i(t_\mu) \; ; \; \mu=1,2,\ldots,N \; ; \; i=1,2,\ldots,n \tag{2-3}$$

$\underline{\alpha}$ is an nx1 column vector consisting of the unknown regression parameters, and $\underline{e}$ is an Nx1 column vector of zero mean Gaussian noise samples.

The LS estimator, $\hat{\underline{\alpha}}_{LS}$, is that function of the observables which minimizes the sum of the squares of the residuals. That is, $\hat{\underline{\alpha}}_{LS}$ minimizes the quadratic form

$$Q(\underline{\alpha}) = \sum_{\mu=1}^{N} \left\{ r_\mu - \sum_{i=1}^{n} \alpha_i g_{i\mu} \right\}^2 \tag{2-4}$$

Hence,

$$\left. \frac{\partial Q(\underline{\alpha})}{\partial \alpha_j} \right|_{\underline{\alpha} = \hat{\underline{\alpha}}_{LS}} = 0 = \sum_{\mu=1}^{N} \left\{ r_\mu - \sum_{i=1}^{n} \hat{\alpha}_{LS_i} \, g_{i\mu} \right\} g_{j\mu} \; ; \; j=1,2,\ldots,n \tag{2-5}$$

The system of equations above is summarized by the equation

$$G^T\underline{R} = G^TG\,\underline{\hat{\alpha}}_{LS} \tag{2-6}$$

which yields the familiar result*

$$\underline{\hat{\alpha}}_{LS} = [G^TG]^{-1}G^T\underline{R} \tag{2-7}$$

It is easy to see that $\underline{\hat{\alpha}}_{LS}$ is an unbiased estimate of $\underline{\alpha}$. That is, let $\underline{\alpha}$ belong to the parameter set A. Then**

$$E[\underline{\hat{\alpha}}_{LS}] = \underline{\alpha}\ \forall\ \underline{\alpha} \in A \tag{2-8}$$

The covariance matrix of any unbiased estimate of $\underline{\alpha}$, say $\underline{\hat{\alpha}}$, (also termed the estimator dispersion matrix) is an nxn matrix of elements, $\{d_{ij};\ i,j = 1,\ldots,n\}$ which we will denote by

$$D(\underline{\hat{\alpha}}) = [d_{ij}] = E[(\underline{\hat{\alpha}} - \underline{\alpha})\ (\underline{\hat{\alpha}} - \underline{\alpha})^T] \tag{2-9}$$

For the LS estimate we have

$$D(\underline{\hat{\alpha}}_{LS}) = [G^TG]^{-1}G^T\Phi\ G[G^TG]^{-1} \tag{2-10}$$

where Φ is the noise covariance matrix,

---

*The existence of $[G^TG]^{-1}$ is assumed. This requires, of course, that $N \geq n$.

**The notation E[...] denotes ensemble expectation and will be used interchangeably with the symbol $\overline{[...]}$.

$$\Phi = [\phi_{\mu\nu}] = E[\underline{e}\,\underline{e}^T] \tag{2-11}$$

Observe that $\hat{\underline{\alpha}}_{LS}$ is linear in the data, and that it in no way involves knowledge of the noise covariance matrix for its construction.

To derive the corresponding formulas for the continuous-time case we may proceed by making use of the Karhunen-Loéve expansion of the random process e(t). Then for a positive definite covariance function, $\phi(t,s)$

$$\phi(t,s) = E[e(t)e(s)] \tag{2-12}$$

$$\lambda_k \psi_k(s) = \int_0^T \phi(t,s)\psi_k(t)\,dt \tag{2-13}$$

$$\int_0^T \psi_k(t)\psi_\ell(t)\,dt = \delta_{k\ell} \tag{2-14}$$

we have

$$r_k = \int_0^T r(t)\psi_k(t)\,dt \tag{2-15}$$

$$g_{ik} = \int_0^T g_i(t)\psi_k(t)\,dt \tag{2-16}$$

$$e_k = \int_0^T e(t)\psi_k(t)\,dt \tag{2-17}$$

where the above equations hold for k = 1,2,....

Hence, we have

$$r_k = \sum_{i=1}^{n} \alpha_i g_{ik} + e_k \; ; \; k = 1,2,... \tag{2-18}$$

or in vector-matrix form

$$\underline{R} = G\,\underline{\alpha} + \underline{e} \tag{2-19}$$

Equation (2-19) is exactly the same as Equation (2-1), except that in (2-19) the dimensions of the vectors and matrices are infinite with respect to the k index. The solution for $\hat{\underline{\alpha}}_{LS}$ obtained in Equation (2-7), therefore, still holds with the following modifications. The nxn matrix G^TG has elements given by

$$(G^TG)_{ij} = \sum_{k=1}^{\infty} g_{ik}g_{kj} = \sum_{k,m=1}^{\infty} \int_0^T g_{ik}g_{mj}\psi_k(t)\psi_m(t)\,dt$$

$$= \int_0^T \left(\sum_{k=1}^{\infty} g_{ik}\psi_k(t)\right)\left(\sum_{m=1}^{\infty} g_{mj}\psi_m(t)\right)\,dt$$

$$= \int_0^T g_i(t)\,g_j(t)\,dt \; ; \quad i,j = 1,2,...,n \tag{2-20}$$

Similarly, the nxl vector $G^T\underline{R}$ has elements given by

$$(G^T\underline{R})_j = \sum_{k=1}^{\infty} g_{jk} r_k = \int_0^T g_j(t)\, r(t)\, dt \; ; \; j = 1,2,\ldots,n \qquad (2\text{-}21)$$

For the estimator dispersion matrix, Equation (2-10) still holds, where (2-20) is used to obtain the matrix $G^T G$, and where the elements of the matrix $G^T \Phi\, G$ are found from

$$(G^T \Phi\, G)_{ij} = \sum_{k,m=1}^{\infty} g_{ik}\, \phi_{km}\, g_{mj} \; ; \; i,j = 1,2,\ldots,n \qquad (2\text{-}22)$$

with

$$\phi_{km} = \int_0^T\int \phi\,(t,s)\psi_k(t)\psi_m(s)\, dt\, ds = \lambda_k\, \delta_{km} \qquad (2\text{-}23)$$

Substituting (2-23) into (2-22) we obtain the desired expression.

$$(G^T \Phi\, G)_{ij} = \int_0^T\int g_i(t)\, \phi\,(t,s)\, g_j(s)\, dt\, ds \; ; \; i,j=1,2,\ldots,n \qquad (2\text{-}24)$$

2.3 Generalized Least Squares Estimation

In the preceding section, the LS estimator for $\underline{\alpha}$ was found by minimizing the sum of the squares of the residuals. The GLS estimator is obtained by minimizing the more general quadratic form

$$Q'(\underline{\alpha}) = \sum_{\mu,\nu=1}^{N} \left\{ r_\mu - \sum_{i=1}^{n} \alpha_i g_{i\mu} \right\} \left\{ r_\nu - \sum_{i=1}^{n} \alpha_i g_{i\nu} \right\} \eta_{\mu\nu} \qquad (2\text{-}25)$$

where $\eta_{\mu\nu}$ are the elements of an arbitrary (real) positive-definite matrix, η. Usually, η is taken to be symmetric. It is apparent that when η is chosen to be the identity matrix, the LS estimator will result. It will be seen shortly that $\hat{\underline{\alpha}}_{GLS}$ is an unbiased estimator of $\underline{\alpha}$ and is linear in the data.

For Gaussian noise with covariance matrix Φ, if η is taken to be Φ^{-1}, then $\hat{\underline{\alpha}}_{GLS}$ is the MVUE of $\underline{\alpha}$. This can be shown in a variety of ways [1, 2].

For the present we shall regard η as any real, symmetric, positive-definite matrix and give the form of $\hat{\underline{\alpha}}_{GLS}$.

Proceeding formally with differentiation as in Equation (2-5) the following matrix equation results.

$$[G^T\eta G]\,\hat{\underline{\alpha}}_{GLS} = G^T\eta\underline{R} \tag{2-26}$$

Hence,

$$\hat{\underline{\alpha}}_{GLS} = [G^T\eta G]^{-1}G^T\eta\underline{R} \tag{2-27}$$

Observe that $\hat{\underline{\alpha}}_{GLS}$ is linear in $\underline{R}$ and is an unbiased estimate of $\underline{\alpha}$, no matter what matrix, η, is used. The dispersion matrix of $\hat{\underline{\alpha}}_{GLS}$ is easily calculated.

$$D(\hat{\underline{\alpha}}_{GLS}) = [G^T\eta G]^{-1}G^T\eta\Phi\eta G[G^T\eta G]^{-1} \tag{2-28}$$

In the case where $\eta = \Phi^{-1}$ this reduces to the well-known result

$$D(\underline{\hat{\alpha}}_o) = [G^T\Phi^{-1}G]^{-1} \qquad (2\text{-}29)$$

where the subscript on $\underline{\hat{\alpha}}_o$ denotes the fact that this is the MVUE of $\underline{\alpha}$. Note from (2-27), with $\eta = \Phi^{-1}$, the explicit dependence of $\underline{\hat{\alpha}}_o$ on the (normalized) noise covariance matrix.

The formulas for the continuous time case, corresponding to Equations (2-20), (2-21) and (2-24), for arbitrary η require special treatment. When the sampled-data form of η is the inverse of a covariance matrix resulting from the sampled form of a covariance function, $\phi(t,t')$, the results of Swerling [7] may be used to find the formulas in the limit as the samples become dense in the observation interval. We will be interested in the dispersion matrix of $\underline{\hat{\alpha}}_o$ when the sampling is dense. In this case

$$(G^T\Phi^{-1}G)_{ij} = \int_0^T g_i(t)\, h_j(t)\, dt \; ; \; i,j = 1,2,\ldots,n \qquad (2\text{-}30)$$

where $h_j(t)$ is the solution to the integral equation

$$\int_0^T h_j(t)\, \phi(t,s)\, dt = g_j(s) \; ; \; s \in [0,T] \; ; \; j = 1,2,\ldots,n \qquad (2\text{-}31)$$

This result is valid whenever $\{h_j(t) \; ; \; j = 1,2,\ldots,n\}$ exist as a solution to the above integral equation and whenever certain other conditions outlined in Swerling [7] and [15] are satisfied.

The estimator dispersion matrix for $\underline{\hat{\alpha}}_o$ is then found using

Equation (2-29).

With the preceding results at our disposal, we are prepared to compare the performance of the LS estimator with that of the MVUE.

2.4 Performance Comparison

The purpose of this section is to disclose a class of problems in which "dramatic improvement" over the performance of the LS estimator would be possible if the noise covariance were known. By dramatic here, we mean at least an order of magnitude reduction in the mean square estimation error.

This investigation is motivated in two ways. Firstly, the comparison provides a measure of the sensitivity of the optimum estimator to imprecise knowledge of the noise covariance. Highly sensitive cases will exhibit dramatic performance differences. Secondly, since a simulation is employed to investigate the performance of the estimation procedure suggested in Chapter IV, it is desirable to simulate examples where the improvement, if any, is not likely to be obscured by simulation innaccuracies. The results of this section provide such examples.

The following treatment will be concerned with the case when $\underline{\alpha}$ consists of only a single parameter, α. The formulas presented in the previous sections can be used for the more general vector parameter case in an obvious manner. Also, the continuous-time formulas will

be emphasized since their use avoids the need for matrix inversions and facilitates hand calculations.

Therefore, for the variance of the LS estimator we have from (2-10), (2-20) and (2-24)

$$V_{LS} = \frac{\int_0^T \int g(t)\, \phi(t,s)\, g(s)\, dt\, ds}{\left[\int_0^T g^2(t)\, dt\right]^2} \tag{2-32}$$

and the minimum variance possible is

$$\overset{o}{V} = \left[\int_0^T h(t)\, \phi\,(t,s)\, h(s)\, dt\, ds\right]^{-1} = \left[\int_0^T h(t)\, g(t)\, dt\right]^{-1} \tag{2-33}$$

where $h(t)$ is the solution to

$$\int_0^T h(t)\, \phi\,(t,s)\, dt = g(s)\ ;\ s\varepsilon[0,T] \tag{2-34}$$

We seek examples for which $\overset{o}{V}$ is much less than V_{LS}. One such example is the following:

Example 1.

$$g(t) = \begin{cases} \cos \omega_o t\ ;\ t\varepsilon[0,T] \\ 0 \qquad\quad \text{otherwise} \end{cases} \tag{2-35}$$

$$\phi\,(t-s) = \sigma^2 \exp[-\beta\ |t-s|\,]^* \tag{2-36}$$

*Note that the noise process is stationary in this example. This will also be the case for Example 2, which follows.

where the parameters ω_o and β are chosen in accordance with the condition that for some non-negative integer, k,

$$0<\beta\ T<<1<\omega_o\ T = \frac{(2k+1)\pi}{2} \tag{2-37}$$

where T is the length of the observation interval. If this is done, then for dense sampling we obtain

$$\frac{\hat{V}}{V_{LS}} \tilde{=} (\beta\ T)\ \frac{(\omega_o T)^2}{(\omega_o T)^2 + 2\beta\ T} \tilde{=} \beta\ T \tag{2-38}$$

To derive the above results we apply Equations (2-32) through (2-34).

$$V_{LS} = \frac{\sigma^2 \int_0^T\int (\cos\omega_o t)\ (\cos\omega_o s)\ \exp(-\beta\ |t-s|)\ dt\ ds}{\left[\int_0^T \cos^2\omega_o t\ dt\right]^2} \tag{2-39}$$

The calculation of the integrals above results in

$$\begin{aligned} V_{LS} = 8\sigma^2 &\{(\omega_o T)^2[1 + (\beta\ T/\omega_o T)^2][1 + (\sin 2\omega_o T)/2\omega_o T]^2\}^{-1} \cdot \\ &\{(\beta\ T/2)\ [1 + (\sin 2\omega_o T)/2\omega_o T] + (\sin^2\ \omega_o T)/2 \\ &+[\exp(-\beta\ T)][(\beta\ T/\omega_o T)^2\cos\omega_o T - (\beta\ T/\omega_o T)\sin\omega_o T] \cdot \\ &[1 + (\beta\ T/\omega_o T)^2]^{-1} - (\beta\ T/\omega_o T)^2[1 + (\beta\ T/\omega_o T)^2]^{-1}\} \end{aligned} \tag{2-40}$$

To calculate $\overset{o}{V}$ we require the solution for h(t) in Equation (2-34)

This solution is given in Swerling [15] when $\phi(t,s)$ is of the form expressed in (2-36) and $g(t)$ is twice differentiable with respect to t for t within the observation interval. When $g(t)$ is given by (2-35) the result is

$$h(t) = \frac{1}{2\sigma^2}\left\{\beta\,(\cos\omega_o t)[1 + (\omega_o T/\beta\,T)^2] + \delta(t) + [\cos\omega_o T - (\omega_o T)(\sin\omega_o T)/\beta\,T]\,\delta(t-T)\right\} \tag{2-41}$$

where $\delta(.)$ is the Dirac delta function.

Using (2-41) in (2-33), we obtain for ϑ

$$\vartheta = \frac{8\,\sigma^2}{2\left[\frac{(\omega_o T)^2}{\beta\,T} + \beta\,T\right] + 4[1 + \cos^2\omega_o T] + \left[\frac{\beta\,T}{\omega_o T} - \frac{\omega_o T}{\beta\,T}\right]\sin 2\omega_o T} \tag{2-42}$$

It can be verified that the quantity ϑ/V_{LS} has local minima when $\omega_o T$ is equal to an odd multiple of $\pi/2$. If $\omega_o T$ and $\beta\,T$ satisfy relation (2-37), the approximate result in (2-38) follows. Note that the dramatic improvement afforded by the optimum estimator, as measured by ϑ/V_{LS}, depends on the ratio of the signal bandwidth to the noise bandwidth; not on the signal-to-noise power ratio per se.

While the above analysis applies strictly for dense sampling, it provides a good indication of the behavior of the time-sampled version of the problem if the sampling rate is sufficiently high. This will be indicated by numerical results later.

Considerable insight can be gained by looking at the above example from the frequency domain point of view. To this end we shall define the signal power spectrum, $S_g(\omega)$, as

$$S_g(\omega) = |\mathcal{F}\{g(t)\}|^2 = \left|\int_{\infty}^{\infty} g(t)\exp(-j\omega t)\, dt\right|^2 \tag{2-43}$$

where $\mathcal{F}\{\ldots\}$ denotes the Fourier transform operation. The noise power spectrum will be defined in the usual fashion by

$$S_e(\omega) = \mathcal{F}\{\phi\ (t-s)\} \tag{2-44}$$

From this it follows that for Example 1

$$S_g(\omega) = \left(\frac{T}{2}\right)^2\left\{\left(\frac{\sin x_1}{x_1}\right)^2 + \left(\frac{\sin x_2}{x_2}\right)^2 + \frac{\cos\omega_o T[\cos\omega_o T - \cos\omega T]}{x_1 x_2}\right\} \tag{2-45}$$

$$S_e(\omega) = \frac{2\beta\ \sigma^2}{\beta^2 + \omega^2} \tag{2-46}$$

where

$$x_1 = \frac{(\omega - \omega_o)\ T}{2} \tag{2-47}$$

$$x_2 = \frac{(\omega + \omega_o)\ T}{2} \tag{2-48}$$

Now it can be observed that when $\omega_o T$ is fixed at an odd multiple of $\pi/2$ the cross-product term (i.e., the last term) in $S_g(\omega)$ vanishes. This permits some separation between the main regions of concentration of signal and noise power.

Figure (2-1) depicts the distribution of signal and noise power in the frequency domain when the parameters are chosen so that dramatic improvement is possible. In this illustration $\omega_o T$ equals $5\pi/2$. (The spectra are symmetric about zero frequency and ω_o.)

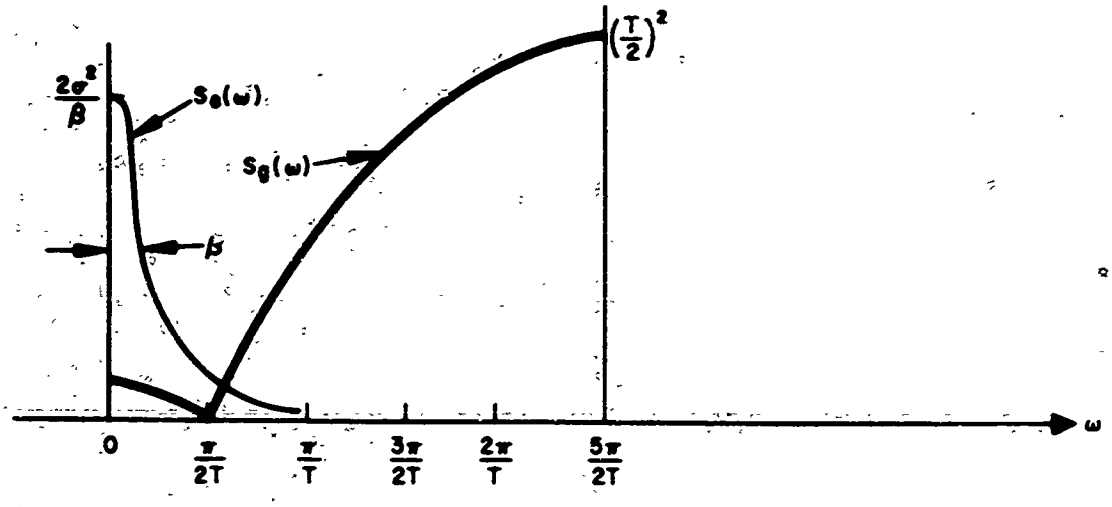

Figure 2-1. Power Spectra for Example 1.

Figure (2-1) suggests that the dual problem, which obtains when the center-frequency locations of the two spectra in Example 1 are interchanged, might also afford an example where dramatic improvement over the performance of the LS estimator is possible. This, in fact, is true.

The specific example is

Example 2.

$$g(t) = \begin{cases} 1 ; & t \in [0,T] \\ 0 & \text{otherwise} \end{cases} \tag{2-49}$$

$$\phi(t-s) = \sigma^2 \exp(-\beta |t-s|)\cos[\omega_c(t-s)] \tag{2-50}$$

When the parameters are chosen so that for some non-negative integer, k,

$$0<\beta T<<1<\omega_c T = (2k+1)\pi \tag{2-51}$$

then corresponding to Equation (2-38) we have

$$\frac{\hat{V}}{V_{LS}} \simeq \left(\frac{\beta T}{2}\right)\left[\frac{\omega_c T[\exp(-\omega_c T)+1]}{\omega_c T[\exp(-\omega_c T)+1] + 2[\exp(-\omega_c T)-1]}\right] \simeq \beta T/2 \tag{2-52}$$

where the last approximation holds when $\omega_c T$ is $>> 1$ (i.e., $k>1$).

The derivation of this result follows.

For V_{LS} we have

$$V_{LS} = \frac{1}{T^2}\iint_0^T \phi(t-s)\,dt\,ds = \frac{\sigma^2}{T^2}\iint_0^T \exp(-\beta|t-s|)\cos[\omega_c(t-s)]dt\,ds$$

$$= \frac{2\sigma^2}{[(\omega_c T)^2 + (\beta T)^2]}\left\{\beta T + \frac{(\omega_c T)^2 - (\beta T)^2}{(\omega_c T)^2 + (\beta T)^2}\left[1 - \exp(-\beta T)\cdot\cos\omega_c T\right]\right.$$

$$\left. - \frac{2\exp(-\beta T)}{(\omega_c T)^2 + (\beta T)^2}(\omega_c T)(\beta T)\sin\omega_c T\right\} \tag{2-53}$$

To calculate $\overset{o}{V}$ it is necessary to solve the integral equation

$$\sigma^2 \int_0^T h(t)\exp(-\beta\,|t-s|)\cos[\omega_c(t-s)]\,dt = 1\ ;\ s\in[0,T] \tag{2-54}$$

for h(t). The details for obtaining this solution are presented in Appendix I. The result is

$$h(t) = K_o\{\exp(-kt) + \exp[+k(t-T)]\} + K_1 + K_2[\delta(t) + \delta(t-T)] \tag{2-55}$$

where

$$k^2 = \omega_c^2 + \beta^2 \tag{2-56}$$

$$K_o = \frac{-k\,\omega_c^2}{\beta\,\sigma^2[(k+\beta) + (k-\beta)\exp(-kT)]} \tag{2-57}$$

$$K_1 = k^2/2\beta\sigma^2 \tag{2-58}$$

$$K_2 = \left[\frac{K_1}{k}\right]\left[\frac{(k+\beta) - (k-\beta)\exp(-kT)}{(k+\beta) + (k-\beta)\exp(-kT)}\right] \tag{2-59}$$

Applying Equation (2-33) we obtain

$$\overset{o}{V} = \left[\frac{2K_o}{k}\,[1-\exp(-kT)] + K_1T + 2K_2\right]^{-1} \tag{2-60}$$

In this example, local minima of $\overset{o}{V}/V_{LS}$ occur when $\omega_c T$ is equal to an add integral multiple of π. The power spectra are displayed in Figure (2-2) with $\omega_c T$ equal to 3π. The formulas for the spectra are:

$$S_g(\omega) = \left(\frac{T}{2}\right)^2\left(\frac{\sin x_3}{x_3}\right)^2 \tag{2-61}$$

$$S_e(\omega) = \sigma^2 \left\{ \frac{\beta}{\beta^2 + (\omega_c - \omega)^2} + \frac{\beta}{\beta^2 + (\omega_c + \omega)^2} \right\} \qquad (2\text{-}62)$$

$$x_3 = \frac{\omega T}{2} \qquad (2\text{-}63)$$

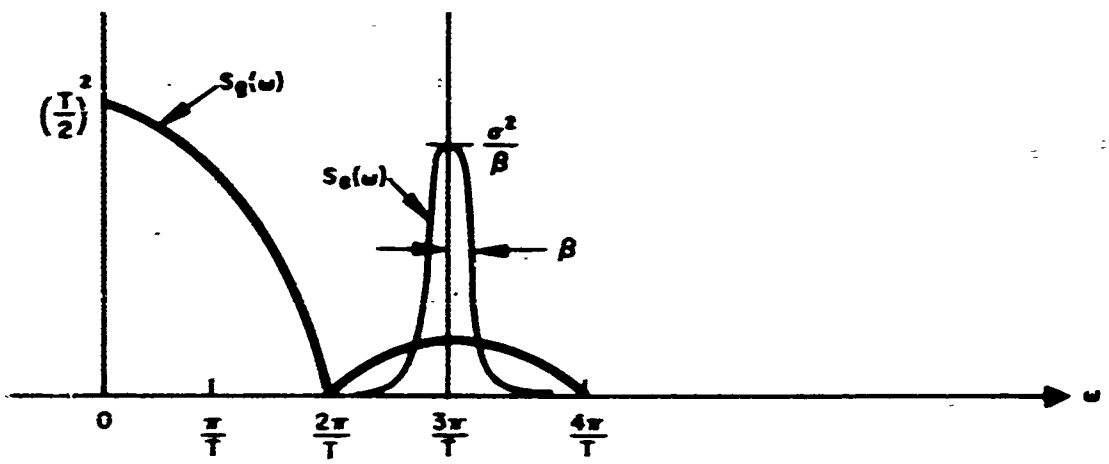

Figure 2-2. Power Spectra for Example 2.

The above examples suggest that a class of problems in which ϑ/V_{LS} is very small (i.e., <<1) is characterized by the property that the noise bandwidth is very small relative to the signal bandwidth, and that the location of the noise "spike" is such that a linear weighting filter operating on the available data can de-emphasize the noise energy

without unduly corrupting the useful signal. It would also appear that this concept could be extended to noise spectra which contain several spikes.

In fact, these statements can be rigorously established whenever the optimum linear unbiased estimator for α can be represented by

$$\hat{\alpha}_o = \int_0^T \overset{o}{w}(t)\, r(t)\, dt \tag{2-64}$$

where $\overset{o}{w}(t)$ denotes the optimum weight function.

By applying the variational calculus to minimize the variance of

$$\hat{\alpha} = \int_0^T w(t)\, r(t)\, dt \tag{2-65}$$

subject to the unbiasedness constraint that

$$\int_0^T w(t)\, g(t)\, dt = 1 \tag{2-66}$$

it is found [36] that $\overset{o}{w}(t)$ must satisfy the integral equation

$$\int_0^T \overset{o}{w}(t)\, \phi\,(t,s)\, dt = \overset{o}{V}\, g(s)\ ;\ s \in [0,T] \tag{2-67}$$

where $\overset{o}{V}$ is the minimum variance obtained, and is given by

$$\overset{o}{V} = \int_0^T \int \overset{o}{w}(t)\, \phi\,(t,s)\, \overset{o}{w}(s)\, dt\, ds \qquad (2\text{-}68)$$

(Multiply both sides of (2-67) by $\overset{o}{w}(t)$ and apply (2-66).)

But $\overset{o}{V}$ is also given by Equations (2-33) and (2-34), since the optimum estimator is unique with probability one [1, 6]. We then have the association between $\overset{o}{w}(t)$ and h(t) given by

$$\overset{o}{w}(t) = \overset{o}{V}\, h(t). \qquad (2\text{-}69)$$

Hence, except for a constant, the optimum weight function is equal to the solution, h(t), of Equation (2-34).

Now for any w(t) in (2-65) and for stationary noise, the variance of the corresponding estimate, $\hat{\alpha}$, is

$$V(\hat{\alpha}) = \int_0^T \int w(t)\, \phi\,(t-s)\, w(s)\, dt\, ds \qquad (2\text{-}70)$$

And if we define k(t) by

$$k(t) = \begin{cases} w(t) & t \in [0,T] \\ 0 & \text{otherwise} \end{cases} \qquad (2\text{-}71)$$

then making use of the convolution theorem [37] and Parseval's formula we have

$$V(\hat{\alpha}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |K(\omega)|^2\, S_e(\omega)\, d\omega \qquad (2\text{-}72)$$

where $K(\omega)$ is the Fourier transform of $k(t)$. Furthermore, the constraint equation may be rewritten as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} K(\omega)\, G^{*}(\omega)\, d\omega = 1 \qquad (2\text{-}73)$$

where $G(\omega)$ is the Fourier transform of $g(t)$.

We now see that $\overset{o}{w}(t)$ is such that the integral in (2-72) is minimized subject to the constraint in (2-73). It is in this fashion that $\overset{o}{w}(t)$ de-emphasizes the noise energy while attempting to preserve the information in the desired signal.

Another point of view which is useful for investigating conditions for dramatic improvement is obtained by employing the Karhunen-Loéve expansion of the noise process. This point of view does not require that the noise process be stationary.

Suppose, in accordance with Equations (2-12) through (2-17), we find the eigenfunctions $\left\{\psi_k(t)\right\}_{k=1}^{\infty}$ and corresponding eigenvalues $\left\{\lambda_k\right\}_{k=1}^{\infty}$ associated with the covariance function, $\phi(t,s)$. Then the power in the k^{th} noise expansion coefficient, e_k, is λ_k, and the power in the k^{th} signal expansion coefficient, g_k, is $\frac{g_k^2}{T}$. Separation of the signal and noise energy distribution occurs if, for every k, the product $\frac{g_k^2 \lambda_k}{T}$ is equal to zero. While total separation cannot

normally be expected, it may be sufficient, in order to obtain $\delta V/V_{LS} << 1$, to have $\frac{g_k^2 \lambda_k}{T} << \max\left\{\frac{(g_k)^2}{T}, \lambda_k\right\}$ whenever either $\frac{(g_k)^2}{T}$ or λ_k is large.

For instance, Van Trees [53] gives the eigenfunctions and eigenvalues for Example 1 (except that the observation interval is taken symmetric about zero for convenience). Hence, for Example 1 with $t \in [-T/2, T/2]$ we have

$$\lambda_k = \frac{2\beta\sigma^2}{\beta^2 + b_k^2} \quad ; k = 1,2,... \tag{2-74}$$

$$\psi_k(t) = \frac{\sqrt{2}}{\sqrt{T}\left[1 + \frac{\sin b_k T}{b_k T}\right]^{1/2}} \cos b_k t \quad ; k \text{ odd} \tag{2-75}$$

$$\psi_k(t) = \frac{\sqrt{2}}{\sqrt{T}\left[1 - \frac{\sin b_k T}{b_k T}\right]^{1/2}} \sin b_k t \quad ; k \text{ even} \tag{2-76}$$

when the b_k are the solutions to the transcendental equation

$$\left[\tan(b_k T/2) + \frac{b_k}{\beta}\right]\left[\tan(b_k T/2) - \frac{\beta}{b_k}\right] = 0 \tag{2-77}$$

If the values of b_k are arranged in increasing order and $\omega_o T$ and βT are chosen in accordance with (2-37) it is found that

$$b_k T \simeq (k - 1)\frac{\pi}{4} \quad ; k \geq 4 \tag{2-78}$$

Meanwhile we have

$$g_k = A_k \int_{-T/2}^{T/2} \cos\omega_o t \cos b_k t \, dt$$

$$= A_k\left(\frac{T}{2}\right)\left\{\frac{\sin[(\omega_o - b_k)T/2]}{(\omega_o - b_k)T/2} + \frac{\sin[(\omega_o + b_k)T/2]}{(\omega_o + b_k)T/2}\right\}, \text{ k odd} \quad (2\text{-}79)$$

$$g_k = A_k \int_{-T/2}^{T/2} \cos\omega_o t \sin b_k t \, dt$$

$$= -A_k\left(T/2\right)\left\{\frac{\cos[(b_k - \omega_o)T/2]}{(b_k - \omega_o)T/2} + \frac{\cos[(b_k + \omega_o)T/2]}{(b_k + \omega_o)T/2}\right\}, \text{k even} \quad (2\text{-}80)$$

where

$$A_k = \begin{cases} \dfrac{\sqrt{2}}{\sqrt{T}\left[1 + \dfrac{\sin b_k T}{b_k T}\right]^{1/2}} & ; \text{ k odd} \\ \dfrac{\sqrt{2}}{\sqrt{T}\left[1 - \dfrac{\cos b_k T}{b_k T}\right]^{1/2}} & ; \text{ k even} \end{cases} \quad (2\text{-}81)$$

Using the above results, it is easily verified that, indeed,

$$\frac{g_k^2 \lambda_k}{T} << \max\left\{\frac{(g_k)^2}{T}, \lambda_k\right\} \text{ for all k.}$$

The above remarks are intended to serve as an intuitive guide to a class of problems where one might expect to be able to construct an

estimator which is significantly better than the LS estimator. There is, in fact, a close analogy between the class of problems mentioned above and problems related to spatial filtering of directional noise sources in radar and sonar applications [21]. In any case, Equations (2-10) and (2-29) may be used to investigate other specific examples.

It is well to point out that for large observation intervals, the class of problems in which the LS estimator can be significantly improved upon is quite small [2, 38]. Even Examples 1 and 2 above are such that $\overset{o}{V}/V_{LS}$ approaches unity as $T \to \infty$. On the other hand, this indicates that many problems in which $D(\hat{\underline{\alpha}}_o) \approx D(\hat{\underline{\alpha}}_{LS})$ for large T, might be more interesting, with regard to sensitivity, when T is small.

Chapter 3

BOUNDS FOR THE JOINT PARAMETER ESTIMATION PROBLEM

3.1 Introduction

Up until now we have been concerned with the performance of the LS estimator for $\underline{\alpha}$ as compared with that of the MVUE, which requires complete knowledge of the normalized covariance function of the noise for its construction. We would like to know the answer to the following question: What is the limit on the accuracy of unbiased estimates of $\underline{\alpha}$ when the normalized covariance function, itself, has unknown parameters? More precisely, what is the greatest lower bound on the variance of unbiased estimates of $\underline{\alpha}$ in the presence of unknown covariance parameters? We are also interested in the form of the estimator which achieves this bound. This, then, is what we will refer to as the joint parameter estimation problem.

Hence, consider the family of Gaussian probability density functions $\{p(\underline{R};\underline{\gamma}) : \underline{\gamma}\varepsilon\Gamma\}$ defined with respect to Lebesgue measure over Euclidean N-space, E_N, where

$$\underline{\gamma}^T = [\underline{\beta}\ \underline{\alpha}]^T \tag{3-1}$$

denotes the (m+n)-dimensional parameter vector composed of the m-dimensional covariance parameter vector, $\underline{\beta}$, and the n-dimensional regression parameter vector, $\underline{\alpha}$, which belong to some parallelopiped, Γ, in E_{m+n}.

The Cramér-Rao (C-R) bound for the joint parameter estimation problem furnishes a lower bound, not only for the variance of unbiased estimates of $\underline{\alpha}$, but also for the mean squared error of estimates for the covariance parameters. These bounds, however, are in general not the greatest lower bounds, and unless they are, they are not attainable by any estimator. Nonetheless, the ease with which the C-R bound can be calculated in many cases, renders it an important bound.

We will calculate the C-R bound for the joint parameter estimation problem in Example 1, assuming that unbiased estimates for all the parameters exist. (An unbiased estimate of the regression parameter always exists; namely the LS estimator. It can also be shown that an unbiased estimate of σ^2 exists. The investigation of whether or not unbiased estimates of the parameters in the normalized covariance function exist when the regression parameter is unknown, is beyond the scope of this work. It is noted, however, that the calculations for the C-R bound which follow can be easily modified to include the effect of a known bias in estimating these parameters.)

The calculation of the C-R bound for Example 1 will reveal that the C-R-efficient* estimates of the regression parameters are uncor-

---

*This terminology refers to those estimates, if they exist, which have the dispersion matrix given by the C-R bound.

related with those of the covariance parameters. Hence, the C-R-efficient regression parameter estimate is unaffected by the lack of knowledge of the covariance parameters. (This conclusion remains valid even when the covariance parameters possess no unbiased estimates.) This is certainly an unsatisfactory bound in view of the fact that the optimum estimator of $\underline{\alpha}$ was found earlier to depend explicitly upon the normalized covariance function, and therefore on β for this example. This motivates an application of Barankin's theory.

The Barankin bound for unbiased estimates of $\underline{\alpha}$ is the greatest lower bound on the mean squared error of such estimates about some pre-chosen parameter point, $\gamma_o \in \Gamma$. The choice of γ_o is completely arbitrary as long as it lies within the allowed parameter space. Whenever at least one unbiased estimator of $\underline{\alpha}$ exists, there is an unbiased estimator of $\underline{\alpha}$ which achieves the Barankin bound. This estimator is termed "locally best" (for $\gamma = \gamma_o$).

In general, only <u>locally</u> best unbiased estimators exist. Rao [1] gives a necessary and sufficient condition for the existence of a uniformly best unbiased estimator (i.e., an MVUE). The theory of complete sufficient statistics [39, 40] offers a sufficient condition for the existence of a MVUE.

The beauty of Barankin's theory lies in the fact that whenever a uniformly best unbiased estimator exists, the Barankin theory will provide us with it, even if no complete sufficient statistic exists.

In our application, we would like to find, if possible, an estimator of $\underline{\alpha}$ which has minimum variance among all unbiased estimates of $\underline{\alpha}$. We must then regard $\underline{\gamma}_o$ as the true value of the parameter $\underline{\gamma}$. If a uniformly best unbiased estimator exists, it will not depend upon the choice of $\underline{\gamma}_o$, and we will have arrived at the MVUE of $\underline{\alpha}$.

Actually, we will show by direct calculation of the Barankin estimator that no MVUE for $\underline{\alpha}$ exists when the normalized covariance function has unknown parameters. This can be shown more expeditiously by making use of the fact that the (locally) best unbiased estimator is unique. Applying this fact and the results of Chapter II which showed that $\hat{\underline{\alpha}}_{GLS} = \hat{\underline{\alpha}}_o$ when $\eta = \phi^{-1}$, and that $\hat{\underline{\alpha}}_{GLS}$ is an unbiased estimate of $\underline{\alpha}$ for arbitrary η, we see that $\hat{\underline{\alpha}}_{GLS}$ is best for $\underline{\alpha}$ if and only if $\eta = \phi^{-1}$. That is, $\hat{\underline{\alpha}}_{GLS}$ with $\eta = \phi^{-1}$ is the Barankin estimator which is best for $\underline{\gamma} = \underline{\gamma}_o$. But this requires the optimum estimator to depend upon $\underline{\beta}_o$; so the Barankin estimator for $\underline{\alpha}$ is only locally best.

Despite the availability of the simple argument above, it is instructive to apply Barankin's theory in detail to the joint estimation problem. In so doing we will develop some of the machinery which is useful in applying Barankin's theory to the estimation of

the covariance parameters.

3.2 Cramer-Rao Bounds for First-Order Autoregressive Gaussian Noise

The problem introduced as Example 1 earlier is actually a first-order autoregressive scheme*. This permits us to write a simple analytic form for the inverse of the noise covariance matrix if the N data samples are uniformly spaced on the interval $[0,T]$†. We shall suppose that this is the case. Then

$$\phi_{\mu\nu} = \sigma^2 \exp[-\beta\,\delta|\mu-\nu|] \;;\; \mu,\nu = 1,\ldots,N \tag{3-2}$$

where

$$\delta = T/(N-1) \tag{3-3}$$

is the interval between time samples.

Now with $\eta = \phi^{-1}$ and

$$\rho = \exp(-\beta\,\delta) \tag{3-4}$$

we have [41]

$$\eta = \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho & & & \\ -\rho & (1+\rho^2) & -\rho & & \\ & -\rho & (1+\rho^2) & -\rho & \\ & & \ddots & \ddots & \ddots \\ & & & -\rho & 1 \end{bmatrix} \tag{3-5}$$

---

*See Section 4.3

†Brennan and Reed [55] show that this noise covariance matrix has a simple inverse even for unequally spaced time samples.

That is, η has only one nonzero off-diagonal on each side of the main diagonal, and the main diagonal entries are $1 + \rho^2$ except for the two corner elements, which are unity.

The joint probability density function of the observables is

$$P(\underline{R};\underline{\gamma}) = \frac{(\det \eta)^{1/2}}{(2\pi)^{N/2}} \cdot \exp\left\{-\frac{1}{2}[\underline{R} - G\underline{\alpha}]^T \eta [\underline{R} - G\underline{\alpha}]\right\} \tag{3-6}$$

(The dependence of η on $\underline{\beta}$ has not been explicitly denoted.)

Now make the following associations between the components of $\underline{\gamma}$ and the unknown parameters $\underline{\beta}$ and α.

$$\underline{\gamma}^T = [\gamma_1\ \gamma_2\ \gamma_3] = [\beta\ \sigma^2\ \alpha] \tag{3-7}$$

and define

$$l_{mn} = -E\left\{\frac{\partial^2 \ln p(\underline{R};\underline{\gamma})}{\partial\gamma_m\ \partial\gamma_n}\right\}\ ;\ m,n = 1,2,3. \tag{3-8}$$

$$L = \left[l_{mn}\right] \tag{3-9}$$

Assuming that all the components of $\underline{\gamma}$ have unbiased estimates, the matrix L^{-1} comprises the desired C-R bound. We shall denote the elements of L^{-1} by l^{mn}. Then, for example, l^{11} is the C-R bound on the variance of unbiased estimates of β. The C-R bound for unbiased estimates of α is given by l^{33}. The covariance of C-R-efficient estimates of α and β is given by l^{13}. Similar remarks hold for the other

elements of L^{-1}.

We now proceed to calculate the elements of L and L^{-1}.

Letting

$$A(\underline{R};\underline{\gamma}) = \ln p(\underline{R};\underline{\gamma}) \tag{3-10}$$

we have

$$A(\underline{R};\underline{\gamma}) = \ln\left[(2\pi)^{-N/2}\right] + \frac{1}{2}\ln[\det(\eta)] - \frac{1}{2}\sum_{\mu,\nu=1}^{N}\eta_{\mu\nu}(r_\mu - \alpha g_\mu)(r_\nu - \alpha g_\nu) \tag{3-11}$$

We may arrive at a simple expression for the determinant of η in this example:

$$\det(\eta) = \sigma^{2N}(1-\rho^2)^{1-N} = \gamma_2^{-N}(1-\rho^2)^{1-N} \tag{3-12}$$

Also, it is convenient to define the quantities

$$Q(\underline{R},\underline{\gamma}) = \sum_{\mu,\nu=1}^{N}\eta_{\mu\nu}(\gamma_1,\gamma_2)\,[r_\mu - \gamma_3 g_\mu]\,[r_\nu - \gamma_3 g_\nu] \tag{3-13}$$

$$A(\gamma_3) = \sum_{\mu=2}^{N-1}[r_\mu - \gamma_3 g_\mu]^2 \tag{3-14}$$

$$B(\gamma_3) = \sum_{\mu=1}^{N-1}[r_\mu - \gamma_3 g_\mu]\,[r_{\mu+1} - \gamma_3 g_{\mu+1}] \tag{3-15}$$

Then for l_{11} we have

$$I_{11} = -E\left\{\frac{\partial^2 A(\underline{R};\underline{Y})}{\partial \gamma_1^2}\right\} = -\frac{1}{2}E\left\{\frac{\partial^2}{\partial\beta^2}\ln[\det(\eta)] - \frac{\partial^2}{\partial\beta^2}Q(\underline{R},\underline{Y})\right\}$$

$$= -\frac{1}{2}E\left\{-\frac{4(1-N)\delta^2\rho^2}{(1-\rho^2)^2} - \frac{4\delta^2\rho^2(1+\rho^2)}{\sigma^2(1-\rho^2)^3}\left[(r_1-\alpha\xi_1)^2 + (r_N-\alpha\xi_N)^2 + 2A(\alpha)\right]\right.$$

$$\left. + \frac{2\delta^2\rho}{\sigma^2(1-\rho^2)^3}\left[(1+6\rho^2+\rho^4)B(\alpha)\right]\right\} \qquad (3\text{-}16)$$

Now observe that

$$\overline{A(\alpha)} = (N-2)\,\sigma^2 \qquad (3\text{-}17)$$

$$\overline{B(\alpha)} = (N-1)\,\sigma^2\rho \qquad (3\text{-}18)$$

$$\overline{(r_1-\alpha\xi_1)^2} = \overline{(r_N-\alpha\xi_N)^2} = \sigma^2 \qquad (3\text{-}19)$$

so that

$$I_{11} = \frac{\delta^2\rho^2(N-1)(3+\rho^2) - 2\delta^2\rho^2(N-1)}{(1-\rho^2)^2} = \frac{\delta^2\rho^2(N-1)(1+\rho^2)}{(1-\rho^2)^2} \qquad (3\text{-}20)$$

The calculation of I_{12} leads to

$$I_{12} = I_{21} = -\frac{1}{2}E\left\{\frac{\partial^2 \ln[\det(\eta)]}{\partial\beta\,\partial(\sigma^2)} - \frac{\partial^2 Q(\underline{R},\underline{Y})}{\partial\beta\,\partial(\sigma^2)}\right\} \qquad (3\text{-}21)$$

The first term above is identically zero since the derivative of the determinant of η with respect to β leads to a function which is independent of σ. The second term is

$$\frac{\partial^2 Q(\underline{R},\underline{\gamma})}{\partial\beta\,\partial(\sigma^2)} = \frac{\partial}{\partial(\sigma^2)}\left\{\frac{-2\delta\rho}{\sigma^2(1-\rho^2)^2}[\rho(r_1-\alpha g_1)^2 + \rho(r_N-\alpha g_N)^2 + 2\rho A(\alpha) - (1+\rho^2)B(\alpha)]\right\}$$

$$= \frac{2\delta\rho}{\sigma^4(1-\rho^2)^2}[\rho(r_1-\alpha g_1)^2 + \rho(r_N-\alpha g_N)^2 + 2\rho A(\alpha) - (1+\rho^2)B(\alpha)] \quad (3\text{-}22)$$

Finally, using (3-17) through (3-19) we obtain

$$I_{12} = \frac{\delta\rho^2(N-1)}{\sigma^2(1-\rho^2)} \quad (3\text{-}23)$$

For I_{13} we have

$$I_{13} = I_{31} = \frac{1}{2}E\left\{\frac{\partial^2 Q(\underline{R},\underline{\gamma})}{\partial\beta\,\partial\alpha}\right\} \quad (3\text{-}24)$$

$$\frac{\partial^2 Q(\underline{R},\underline{\gamma})}{\partial\beta\,\partial\alpha} = \frac{\partial}{\partial\alpha}\left\{\frac{2\delta\rho}{\sigma^2(1-\rho^2)}[B(\alpha) - \rho A(\alpha)] - \frac{2\delta\rho^2}{\sigma^2(1-\rho^2)^2}[(r_1-\alpha g_1)^2 + (r_N-\alpha g_N)^2 + (1+\rho^2)A(\alpha) - 2\rho B(\alpha)]\right\}$$

$$\frac{\partial^2 Q(\underline{R},\underline{\gamma})}{\partial\beta\,\partial\alpha} = \frac{2}{\sigma^2(1-\rho^2)}\left\{2\delta\rho^2\sum_{\mu=2}^{N-1} e_\mu g_\mu - \delta\rho\sum_{\mu=1}^{N-1}[e_\mu g_\mu + e_{\mu+1}g_{\mu+1}]\right\}$$

$$+\frac{4\delta\rho^2}{\sigma^2(1-\rho^2)^2}\left\{e_1 g_1 + e_N g_N + (1+\rho^2)\sum_{\mu=2}^{N-1} e_\mu g_\mu\right.$$

$$\left.-\rho\sum_{\mu=1}^{N-1}[e_\mu g_\mu + e_{\mu+1}g_{\mu+1}]\right\} \qquad (3\text{-}25)$$

And since the noise is zero mean

$$l_{13} = l_{31} \equiv 0 \qquad (3\text{-}26)$$

For l_{22} we obtain

$$l_{22} = -\frac{1}{2}E\left\{\frac{\partial^2 \ln[\det(\eta)]}{\partial^2(\sigma^2)} - \frac{\partial^2 Q(\underline{R},\underline{\gamma})}{\partial^2(\sigma^2)}\right\} \qquad (3\text{-}27)$$

$$\frac{\partial^2 \ln[\det(\eta)]}{\partial^2(\sigma^2)} = \frac{\partial}{\partial(\sigma^2)}\left\{\frac{\partial}{\partial(\sigma^2)}[-N\ln(\sigma^2) + (1-N)\ln(1-\rho^2)]\right\}$$

$$= \frac{\partial}{\partial(\sigma^2)}[-N/\sigma^2] = \frac{N}{\sigma^4} \qquad (3\text{-}28)$$

$$\frac{\partial^2 Q(\underline{R},\underline{\gamma})}{\partial^2(\sigma^2)} = \frac{2}{\sigma^6(1-\rho^2)}[(r_1 - \alpha g_1)^2 + (r_N - \alpha g_N)^2$$

$$+ (1+\rho^2)A(\alpha) - 2\rho B(\alpha)] \qquad (3\text{-}29)$$

Hence,

$$l_{22} = \frac{-N}{2\sigma^4} + \frac{N}{\sigma^4} = \frac{N}{2\sigma^4} \tag{3-30}$$

We find l_{23} from

$$l_{23} = l_{32} = -E\left\{\frac{\partial^2 \Lambda(\underline{R};\underline{\gamma})}{\partial(\sigma^2)\,\partial\alpha}\right\} = \frac{1}{2}\overline{\frac{\partial^2 Q(\underline{R},\underline{\gamma})}{\partial(\sigma^2)\partial\alpha}} \tag{3-31}$$

$$\frac{\partial^2 Q(\underline{R},\underline{\gamma})}{\partial\sigma^2\,\partial\alpha} = \frac{\partial}{\partial\alpha}\left\{\frac{-1}{\sigma^4(1-\rho^2)}\left[(r_1 - \alpha g_1)^2 + (r_N - \alpha g_N)^2 + (1+\rho^2)A(\alpha) - 2\rho B(\alpha)\right]\right\}$$

$$= \frac{2}{\sigma^4(1-\rho^2)}\left[e_1 g_1 + e_N g_N + (1+\rho^2)\sum_{\mu=2}^{N-1} e_\mu g_\mu - \rho\sum_{\mu=1}^{N-1}(e_\mu g_\mu + e_{\mu+1} g_{\mu+1})\right] \tag{3-32}$$

That is,

$$l_{23} = l_{32} \equiv 0 \tag{3-33}$$

The remaining element of L is l_{33}.

$$l_{33} = -E\left\{\frac{\partial^2 \Lambda(\underline{R};\underline{\gamma})}{\partial\alpha^2}\right\} = \frac{1}{2}E\left\{\frac{\partial^2 Q(\underline{R},\underline{\gamma})}{\partial\alpha^2}\right\} \tag{3-34}$$

$$\frac{\partial^2 Q(\underline{R},\underline{Y})}{\partial \alpha^2} = \frac{\partial}{\partial \alpha}\left\{\frac{-2}{\sigma^2(1-\rho^2)}\left[(r_1 - \alpha g_1)g_1 + (r_N - \alpha g_N)g_N\right.\right.$$

$$+ (1+\rho^2)\sum_{\mu=2}^{N-1}(r_\mu - \alpha g_\mu)g_\mu - \rho\sum_{\mu=1}^{N-1}(r_\mu - \alpha g_\mu)g_\mu$$

$$\left.\left. - \rho\sum_{\mu=1}^{N-1}(r_{\mu+1} - \alpha g_{\mu+1})g_{\mu+1}\right]\right\}$$

$$= \frac{2}{\sigma^2(1-\rho^2)}\left[g_1^2 + g_N^2 + (1+\rho^2)\sum_{\mu=2}^{N-1} g_\mu^2\right.$$

$$\left. - \rho\sum_{\mu=1}^{N-1} g_\mu^2 - \rho\sum_{\mu=1}^{N-1} g_{\mu+1}^2\right]$$

$$= \frac{2(1-\rho)}{\sigma^2(1-\rho^2)}\left[g_1^2 + g_N^2 + (1-\rho)\sum_{\mu=2}^{N-1} g_\mu^2\right] \qquad (3\text{-}35)$$

Hence,

$$l_{33} = \frac{1}{\sigma^2(1+\rho)}\left[g_1^2 + g_N^2 + (1-\rho)\sum_{\mu=2}^{N-1} g_\mu^2\right] \qquad (3\text{-}36)$$

To summarize the preceding results, we shall write the complete L matrix.

$$L = \begin{bmatrix} \frac{(N-1)(1+\rho^2)\delta^2\rho^2}{(1-\rho^2)^2} & \frac{(N-1)\delta\rho^2}{\sigma^2(1-\rho^2)} & 0 \\ \frac{(N-1)\delta\rho^2}{\sigma^2(1-\rho^2)} & \frac{N}{2\sigma^4} & 0 \\ 0 & 0 & \frac{g_1^2+g_N^2+(1-\rho)\sum_{\mu=2}^{N-1} g_\mu^2}{\sigma^2(1+\rho)} \end{bmatrix} \quad (3\text{-}37)$$

It is now apparent from the locations of the zeroes in this matrix that the C-R bounds on the estimates of the covariance parameters are not affected by the presence of the unknown regression parameter, α. Similarly, the bound on the regression parameter estimate is unaffected by the presence of the unknown covariance parameters.

Furthermore, though the above analysis was made assuming a single unknown linear regression parameter, the results pertaining to the covariance parameters are unchanged for an arbitrary regression function, $f(t,\underline{\alpha})$. This follows from the fact that the bounds pertinent to the covariance parameters are insensitive to the presence of the regression parameters as long as zeroes appear in all the elements of the L matrix which involve a mixture of covariance and regression parameters (for example; l_{13} and l_{23} above). An inspection of the previous calculations for the quantities l_{13} and l_{23} reveals that this will be the case for arbitrary $f(t,\underline{\alpha})$ by virtue of the fact that

$Q(\underline{R},\underline{Y})$ is a quadratic form in $\underline{R}$.

The actual bounds are found from L^{-1}. If we let $\hat{\beta}$ and $\hat{\sigma}^2$ denote unbiased estimates of their respective true values, β_o and σ_o^2, then the C-R bounds for the covariance parameters are

$$\overline{(\hat{\beta} - \beta_o)^2} \geq \frac{N(1-\rho_o^2)^2}{(N-1)\delta^2\rho_o^2[N(1-\rho_o^2)+2\rho_o^2]} \tag{3-38}$$

$$\overline{(\hat{\sigma}^2 - \sigma_o^2)^2} \geq \frac{2\sigma_o^4(1+\rho_o^2)}{N(1-\rho_o^2) + 2\rho_o^2} \tag{3-39}$$

where ρ_o is given by (3-4) with $\beta = \beta_o$.

The covariance between C-R-efficient estimates of β and σ^2 is given by

$$1^{12} = \frac{2\sigma_o^2(1 - \rho_o^2)}{\delta[N(\rho_o^2 - 1) - 2\rho_o^2]} \tag{3-40}$$

It is interesting to observe that 1^{12}, in general, is not equal to zero. This indicates that the C-R bound for estimating β is increased when the noise power, σ^2, is unknown.

The appropriate results for dense sampling on an interval of length T may be obtained from the above expressions by taking the limit as $N \to \infty$ and $\delta \to 0$ in such a way that $(N-1)\delta = T$. This yields

$$\overline{(\hat{\beta} - \beta_o)^2} \geq \frac{2\beta_o^2}{\beta_o T + 1} \tag{3-41}$$

$$\overline{(\hat{\sigma}^2 - \sigma_o^2)^2} \geq \frac{2\sigma_o^4}{\beta_o T + 1} \tag{3-42}$$

$$1^{12} = \frac{-2\beta_o\sigma_o^2}{\beta_o T + 1} \tag{3-43}$$

In the case when $f(t,\underline{\alpha})$ involves only a single linear regression parameter, the C-R bound on unbiased estimates for α is given by the reciprocal of 1_{33}. It can easily be verified that this agrees with the result predicted by Equation (2-29). Thus, the C-R bound for the regression parameter estimate is achievable, and is attained by the MVUE discussed in Section 2.3. This conclusion also holds for the vector $\underline{\alpha}$ case.

But since the MVUE for $\underline{\alpha}$ in this example requires knowledge of β, we cannot expect to achieve the C-R bound for estimates of $\underline{\alpha}$ when β is unknown and the observation interval is finite. Indeed, an investigation of the Barankin bound, which follows, will reveal that the C-R bound on unbiased estimates of $\underline{\alpha}$ can be attained only if all the parameters in the normalized noise covariance function are known precisely.

3-3 Barankin Bound

In this section we shall calculate the Barankin bound for estimation of the n-dimensional parameter $\underline{\alpha}$ when the covariance function has unknown parameters $\underline{\beta}$. We shall not need to specify explicitly what the parametric dependence on $\underline{\beta}$ is in order to obtain our results. This lends considerable generality to the analysis.

In the process of calculating the Barankin bound for unbiased estimates of $\underline{\alpha}$ we will develop some of the machinery which could prove useful in constructing estimates of the unknown covariance parameters.

The reader is referred to Appendix II for a presentation of the theory pertinent to this section and for elaboration on the essential concepts. We will repeat the basic formulas here for convenience.

Let $\underline{\gamma}$ denote the joint vector of unknown parameters belonging to the parameter space Γ as in (3-1). The function

$$G(\underline{\gamma},\underline{\gamma}'|\underline{\gamma}_o) = \int_{E_N} \frac{p(\underline{R};\underline{\gamma})\ p(\underline{R};\underline{\gamma}')}{p(\underline{R};\underline{\gamma}_o)}\ d\underline{R}\ ;\ \underline{\gamma},\underline{\gamma}',\underline{\gamma}_o \in \Gamma \qquad (3\text{-}44)$$

is central to the technique outlined in Appendix II for calculating the Barankin bound. Here $\underline{\gamma}_o$ is an arbitrary fixed point in Γ, and is the point about which the Barankin estimator has minimum mean squared error. We shall let $\underline{\gamma}_o$ correspond to the true value of the unknown

parameter $\underline{\gamma}$ in an attempt to find a MVUE for each component of $\underline{\alpha}$.

It is required that $G(\underline{\gamma},\underline{\gamma}'|\underline{\gamma}_o)$ be bounded for all values of $\underline{\gamma}\in\Gamma$ whenever $\underline{\gamma}'$ is in a region (in general depending upon $\underline{\gamma}_o$) $\Gamma_{\underline{\gamma}_o}\subset\Gamma$. Once the density function $p(\underline{R};\underline{\gamma})$ is specified, $G(\underline{\gamma},\underline{\gamma}'|\underline{\gamma}_o)$ can be calculated, and the region $\Gamma_{\underline{\gamma}_o}$ can be found.

We then consider the n-dimensional vector of functions or generalized functions, $d\underline{\lambda}(\underline{\gamma})$, defined over Γ for which the integrals

$$\int_{\Gamma_{\underline{\gamma}_o}} G(\underline{\gamma},\underline{\gamma}'|\underline{\gamma}_o)\, d\underline{\lambda}(\underline{\gamma}')$$

$$\int_{E_N}\int_{\Gamma_{\underline{\gamma}_o}} \frac{p(\underline{R};\underline{\gamma}')}{p(\underline{R};\underline{\gamma}_o)}\, d\underline{\lambda}(\underline{\gamma}')\, p(\underline{R};\underline{\gamma})\, d\underline{R}$$

exist and are equal. Suppose we can find such a vector, say $d\underline{\lambda}^o(\underline{\gamma}) = [d\lambda_1^o(\underline{\gamma}) \ldots\ldots d\lambda_n^o(\underline{\gamma})]$, for which

$$\int_{\Gamma_{\underline{\gamma}_o}} G(\underline{\gamma},\underline{\gamma}'|\underline{\gamma}_o)\, d\underline{\lambda}^o(\underline{\gamma}') \equiv \underline{\alpha} - \underline{\alpha}_o \;;\; \forall\, \underline{\gamma}\in\Gamma \qquad (3\text{-}5)$$

and let $\sigma_B^2(\alpha_{oi})$ denote the Barankin bound for the estimation of α_i, the i^{th} component of $\underline{\alpha}$. Then

$$\sigma_{\beta}^{2}(\alpha_{oi}) = \int_{\Gamma_{Y_o}} (\alpha_i - \alpha_{oi})\, d\lambda_i^{o}(\underline{Y}) \quad ; i = 1,\ldots,n \tag{3-45}$$

The estimator which attains this bound will be denoted by $\hat{\underline{\alpha}}_{\beta}$, and $\hat{\alpha}_{\beta_i}$ is given by

$$\hat{\alpha}_{\beta_i} = \int_{\Gamma_{Y_o}} \frac{p(\underline{R};\underline{Y})}{p(\underline{R};\underline{Y}_o)}\, d\lambda_i^{o}(\underline{Y}) + \alpha_{oi} \quad ; i = 1,\ldots,n \tag{3-47}$$

We will now proceed to calculate $G(\underline{Y},\underline{Y}'|\underline{Y}_o)$ when $p(\underline{R};\underline{Y})$ is as expressed in (3-6). Here we will denote explicitly the dependance of η upon the parameter vector $\underline{\beta}$ via the notation η_{β}.

We have

$$G(\underline{Y},\underline{Y}'|\underline{Y}_o) = \frac{[\det(\eta_{\beta})\det(\eta_{\beta'})]^{1/2}}{(2\pi)^{N/2}[\det(\eta_{\beta_o})]^{1/2}} \times$$

$$\int_{E_N} \exp -\frac{1}{2}\{[\underline{R} - G\underline{\alpha}]^{T}\eta_{\beta}[\underline{R} - G\underline{\alpha}]$$

$$+ [\underline{R} - G\underline{\alpha}']^{T}\eta_{\beta'}[\underline{R} - G\underline{\alpha}']$$

$$- [\underline{R} - G\underline{\alpha}_o]^{T}\eta_{\beta_o}[\underline{R} - G\underline{\alpha}_o]\}\ d\underline{R} \tag{3-48}$$

Now define the following matrices:

$$H = [\eta_\beta + \eta_{\beta'} - \eta_{\beta_o}] \tag{3-49}$$

$$B = [G\underline{\alpha}]^T \eta_\beta + [G\underline{\alpha}']^T \eta_{\beta'} - [G\underline{\alpha}_o]^T \eta_{\beta_o} \tag{3-50}$$

$$C = [G\underline{\alpha}]^T \eta_\beta [G\underline{\alpha}] + [G\underline{\alpha}']^T \eta_{\beta'} [G\underline{\alpha}'] - [G\underline{\alpha}_o]^T \eta_{\beta_o} [G\underline{\alpha}_o] \tag{3-51}$$

Then,

$$G(\underline{Y},\underline{Y}'|\underline{Y}_o) = \frac{(\det\eta_\beta)^{1/2}(\det\eta_{\beta'})^{1/2} \exp(-C/2)}{(2\pi)^{N/2} (\det\eta_{\beta_o})^{1/2}} \quad X$$

$$\int_{E_N} \exp -\frac{1}{2} [\underline{R}^T \underline{H} \underline{R} - \underline{R}^T \underline{B}^T - \underline{B}\underline{R}] \, d\underline{R} \tag{3-52}$$

Assume that H^{-1} exists, and make the transformation

$$\hat{\underline{R}}^T = \underline{R}^T H^{1/2} \tag{3-53}$$

in the above integral. The integral then becomes

$$I = \int_{E_N} \exp -\tfrac{1}{2}[\hat{\underline{R}}^T\hat{\underline{R}} - \hat{\underline{R}}^T H^{-\frac{1}{2}} \underline{B}^T - \underline{B} H^{-\frac{1}{2}} \hat{\underline{R}}] \cdot [\det H]^{-\frac{1}{2}} \, d\hat{\underline{R}}$$

$$= [\det H]^{-\frac{1}{2}} \int_{E_N} \exp -\frac{1}{2}\left\{\left[\hat{\underline{R}}^T - \underline{B} H^{-1/2}\right]\left[\hat{\underline{R}} - \left(\underline{B} H^{-1/2}\right)^T\right] - \underline{B} H^{-1} \underline{B}^T\right\} d\hat{\underline{R}}$$

$$= [\det H]^{-\frac{1}{2}} \exp[\tfrac{1}{2} (\underline{B} H^{-1} \underline{B}^T)] \cdot (2\pi)^{N/2} \tag{3-54}$$

Therefore,

$$G(\underline{\gamma},\underline{\gamma}'|\underline{\gamma}_o) = \frac{(\det\eta_{\beta'})^{1/2}(\det\eta_{\beta'})^{1/2}\exp[-\frac{1}{2}(C - \underline{E}\underline{H}^{-1}\underline{E}^T)]}{(\det H)^{1/2}\ (\det \eta_{\beta_o})^{1/2}} \qquad (3\text{-}55)$$

This expression is valid for all values of $\underline{\gamma}$ and whenever β' lies in a region, Γ_{β_o} (depending only upon β_o), which ensures the existence of H^{-1}.

Now consider the generalized function

$$d\tilde{\lambda}_i(\underline{\gamma}) = \delta(\underline{\beta} - \underline{\beta}_o)\ \delta'(\alpha_i - \alpha_{oi})\ \delta(\underset{\sim}{\alpha}_{\tilde{i}} - \underset{\sim}{\alpha}_{\tilde{i}o})\ d\underline{\gamma}\ ;\ i = 1,\ldots,n \qquad (3\text{-}56)$$

where $\underset{\sim}{\alpha}_{\tilde{i}}$ denotes the vector $\underline{\alpha}$ with the $i^{\underline{th}}$ component deleted, $\delta(.)$ denotes the Dirac distribution, and $\delta'(.)$ denotes the first derivative of $\delta(.)$ [37].

It can be verified that this is a legitimate choice of $d\underline{\lambda}$ in terms of the existence and equality of the integrals preceding Equation (3-45). It can also be verified that

$$\int_{\Gamma_{\beta_o}} G(\underline{\gamma},\underline{\gamma}'|\underline{\gamma}_o)\ d\tilde{\lambda}_i(\underline{\gamma}') = \sum_{j=1}^{n} (G^T\eta_{\beta_o}G)_{ij}\ (\alpha_j - \alpha_{oj}) \qquad (3\text{-}57)$$

and (3-57) holds for $i = 1,\ldots,n$. In vector-matrix form we have

$$\int_{\Gamma_{\beta_o}} G(\underline{\gamma},\underline{\gamma}'|\underline{\gamma}_o)\ d\underline{\tilde{\lambda}}(\underline{\gamma}') = [G^T\eta_{\beta_o}G][\underline{\alpha} - \underline{\alpha}_o] \qquad (3\text{-}58)$$

Therefore, the desired choice for $d\underline{\lambda}^{o}$ satisfying (3-55) is

$$d\underline{\lambda}^{o}(\underline{\gamma}) = [G^T \eta_{\beta_c} G]^{-1} \, d\underline{\tilde{\lambda}}(\underline{\gamma}) \tag{3-59}$$

The Barankin bound for estimating the $i^{\underline{th}}$ component of $\underline{\alpha}_o$ is:

$$\sigma_B^2(\alpha_{oi}) = \int_{\Gamma_{\beta_o}} (\alpha_i - \alpha_{oi}) \sum_{j=1}^{n} (G^T \eta_{\beta_o} G)^{ij} \, d\tilde{\lambda}_j(\underline{\gamma}) \tag{3-60}$$

where $(G^T \eta_{\beta_o} G)^{ij}$ denotes the $ij^{\underline{th}}$ element of the matrix $[G^T \eta_{\beta_o} G]^{-1}$.

Using (3-56) we find that

$$\sigma_B^2(\alpha_{oi}) = [G^T \eta_{\beta_c} G]^{ii} \tag{3-61}$$

This agrees exactly with the result in Equation (2-29). Hence, as mentioned earlier, the uniqueness of the locally best unbiased estimator implies that the Barankin estimator for $\underline{\alpha}$, which is best about the true value of $\underline{\gamma}$, is equal to the GLS estimator with η_{β_o} as the weight matrix. This can also be verified by direct calculation using Equation (3-47). Hence, we have

$$\begin{aligned} \hat{\underline{\alpha}}_B &= \underline{\alpha}_o + [G^T \eta_{\beta_o} G]^{-1} \int_{\Gamma_{\beta_o}} \exp -\frac{1}{2} \Big\{ [\underline{R}-G\underline{\alpha}]^T \eta_\beta [\underline{R}-G\underline{\alpha}] \\ &\qquad - [\underline{R}-G\underline{\alpha}_o]^T \eta_{\beta_o} [\underline{R}-G\underline{\alpha}_o] \Big\} \, d\underline{\tilde{\lambda}}(\underline{\gamma}) \\ &= \underline{\alpha}_o + [G^T \eta_{\beta_o} G]^{-1} G^T \eta_{\beta_o} [\underline{R} - G\underline{\alpha}_o] \\ &= [G^T \eta_{\beta_o} G]^{-1} G^T \eta_{\beta_o} \underline{R} \end{aligned} \tag{3-62}$$

3.4 Implications of the Results Obtained so far

We have shown that there is no MVUE for $\underline{\alpha}$ when the normalized noise covariance function has unknown parameters, $\underline{\beta}$. However, we have also seen that in some cases there is a dramatic difference between the performance of the MVUE of $\underline{\alpha}$ and the LS estimate of $\underline{\alpha}$. It is natural, then, to ask if, in the case of unknown covariance parameters, there is some estimate of $\underline{\alpha}$ which is significantly better than the LS estimate, though naturally not as good as the MVUE. Unfortunately, it is difficult to lend structure to the problem of searching for such an estimate.

One approach, however, is to construct the Barankin estimator for $\underline{\alpha}$ which is best about an estimate of the true value of $\underline{\beta}$. This really amounts to using the GLS estimator for $\underline{\alpha}$ with an estimate of $\underline{\beta}$ used to construct the appropriate weight matrix. This will not necessarily lead to an unbiased estimate of $\underline{\alpha}$; unless the estimate of $\underline{\beta}$ is made from data which is statistically independent from the data used for estimating $\underline{\alpha}$. Nonetheless, it is possible that the bias will be small enough to be acceptable; especially if the mean squared error in estimating $\underline{\alpha}$ is significantly reduced. Moreover, in some cases this procedure will provide an unbiased estimate of $\underline{\alpha}$, even if the estimate of $\underline{\beta}$ is correlated with the estimate of $\underline{\alpha}$. (This will be elaborated upon in the next chapter.)

This leaves us with the problem of estimating the parameter $\underline{\beta}$. We could use the maximum likelihood (ML) estimate of $\underline{\beta}$ if it is convenient to do so. (Such a procedure would actually amount to joint-ML estimation of $\underline{\alpha}$ and $\underline{\beta}$, since the noise is Gaussian.) Or we could employ Baram's theory to estimate $\underline{\beta}$. For example, we could perform a dissection of the parameter space for $\underline{\beta}$ and apply the technique in Appendix II to obtain an unbiased estimate of each point in the dissection. For this we need only specify the parametric dependence of η on $\underline{\beta}$ and use (3-55) for $G(\underline{Y},\underline{Y}'|\underline{Y}_0)$.

In principle either of these techniques, or others not yet mentioned, could be used, though one method may be more practical than the others in a specific case. In the next chapter we shall consider an estimate of $\underline{\beta}$ (where $\underline{\beta}$ denotes the unknown parameters in the normalized noise covariance function) based upon the known functional form of the normalized noise covariance function, $\theta(\tau,\underline{\beta}) = \dfrac{\phi(\tau,\underline{\beta})}{[e(t)]^2}$. This estimate proves to be remarkably effective for improving the IS estimate of $\underline{\alpha}$ in Examples 1 and 2.

Chapter 4

THE ITERATIVE ESTIMATION PROCEDURE

4.1 Introduction

In this chapter we will explore a particular method of estimating the pertinent covariance parameters in Examples 1 and 2. We will then incorporate the covariance parameter estimates into an estimate of the regression parameter. Though developed thoroughly only for Examples 1 and 2, the method is immediately extendable to other problems in which the normalized noise covariance function contains unknown nonrandom parameters.

Our method of estimating the unknown covariance parameters is based upon knowing the functional form, $\theta(\tau,\underline{\beta}) = \phi(\tau,\underline{\beta})/\phi(0,\underline{\beta})$, of the normalized noise covariance function.

We will use the LS estimate of $\underline{\alpha}$ to get an initial estimate of the regression function. After subtracting the estimated regression function from the data, we have an estimate of the noise process. This estimate allows us to estimate $\underline{\beta}$ using the function $\theta(\tau,\underline{\beta})$. We then generate the estimated normalized covariance function $\hat{\theta}(\tau) = \theta(\tau,\hat{\underline{\beta}})$, and the corresponding matrix $\hat{\theta}$. Finally, we construct the GLS estimate of $\underline{\alpha}$ by using $\hat{\eta} = \hat{\theta}^{-1}$ for the weight matrix. At this point we repeat the process again, this time using the previously con-

structed GLS estimate of $\underline{\alpha}$ to estimate the noise process and then $\underline{\beta}$. This revised estimate of $\underline{\beta}$ leads to a revised estimate of $\underline{\alpha}$. This is the iterative estimation procedure.

We will investigate the bias and the mean squared error of the iterative estimate of $\underline{\alpha}$. The bias can be investigated analytically to some extent. The mean squared error will be investigated by citing experimental results of simulating Examples 1 and 2 on a digital computer.

As a prelude to the development of the desired estimation procedure for $\underline{\beta}$, and for the sake of completeness, we include a brief discussion of the joint maximum likelihood (ML) estimator for $\underline{\gamma}$. It will be seen that this estimator is not practical for Examples 1 and 2. We also present an estimate of $\underline{\gamma}$ which is really only applicable to autoregressive schemes, but which serves to motivate the procedure we employ to estimate $\underline{\beta}$ in Examples 1 and 2.

4.2 Maximum Likelihood Estimation

The ML estimate of $\underline{\gamma}$ is defined as the solution to the likelihood equation

$$\left.\frac{\partial \Lambda(\underline{R};\underline{\gamma})}{\partial \underline{\gamma}}\right|_{\underline{\gamma}=\hat{\underline{\gamma}}} = \underline{0} \qquad (4\text{-}1)$$

where $\Lambda(\underline{R};\underline{\gamma})$ is as in (3-10) and $\hat{\underline{\gamma}}$ denotes the ML estimate of $\underline{\gamma}$.

For Example 1, this leads to the system of equations

$$\left.\frac{\partial L(\underline{\beta},\underline{\gamma})}{\partial \alpha}\right|_{\underline{\gamma}=\hat{\underline{\gamma}}} = \sum_{\mu,\nu=1}^{N} \hat{\eta}_{\mu\nu}(r_\mu - \alpha\xi_\mu)\xi_\nu = 0 \qquad (4\text{-}2)$$

$$\left.\frac{\partial L(\underline{\beta},\underline{\gamma})}{\partial \beta}\right|_{\underline{\gamma}=\hat{\underline{\gamma}}} = \frac{2\hat{\sigma}^2\alpha(1-N)}{(1-\hat{\rho}^2)} - \sum_{\mu,\nu=1}^{N} \hat{\eta}'_{\mu\nu}(r_\mu - \alpha\xi_\mu)(r_\nu - \alpha\xi_\nu) = 0 \qquad (4\text{-}3)$$

where $\hat{\rho}$ is given by (3-4) with $\beta = \hat{\beta}$, and $\hat{\eta}'_{\mu\nu} = \left.\frac{\partial \eta_{\mu\nu}}{\partial \beta}\right|_{\beta=\hat{\beta}}$. Using the result in Equation (3-5) we may rewrite Equation (4-2) as

$$\hat{\alpha} = \frac{\sum_{\mu=2}^{N}(r_\mu - \hat{\rho}r_{\mu-1})(\xi_\mu - \hat{\rho}\xi_{\mu-1}) + r_1\xi_1(1-\hat{\rho}^2)}{\sum_{\mu=2}^{N}(\xi_\mu - \hat{\rho}\xi_{\mu-1})^2 + \xi_1^2(1-\hat{\rho}^2)} \qquad (4\text{-}4)$$

This system of equations must be solved simultaneously for $\hat{\alpha}$ and $\hat{\rho}$ to obtain the ML estimate of $\underline{\gamma}$. This is indeed a difficult task, even on a computer; and the sibuation is worse for Example 2.

4.3 An Approach for Autoregressive Schemes

Suppose we have data of the form

$$r_\mu = \alpha\xi_\mu + e_\mu \; ; \; \mu = 1,\ldots,N \qquad (4\text{-}5)$$

where, for all μ

$$e_\mu = \rho e_{\mu-1} + w_\mu \qquad (4\text{-}6)$$

$$\overline{w_\mu} = 0 \tag{4-7}$$

$$\overline{w_\mu w_\nu} = \sigma_w^2 \delta_{\mu\nu} \tag{4-8}$$

Here w_μ denotes a sample of a zero mean, white, Gaussian noise process, and the e_μ are samples of a Gaussian first order autoregressive scheme. Let e_o denote e_μ at the initial time, t_o. As $t_o \to -\infty$, a stable solution for the difference Equation (4-6) exists if $|\rho|<1$. This solution is [42]

$$e_\mu = \sum_{k=0}^{\infty} \rho^k w_{\mu-k} \tag{4-9}$$

Observe that the variance of the $\{e_\mu\}$ process is

$$\sigma^2 = \overline{e_\mu^2} = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \rho^k \rho^l \overline{w_{\mu-k} w_{\mu-l}} = \sum_{k=0}^{\infty} \sigma_w^2 \rho^{2k} = \frac{\sigma_w^2}{(1-\rho^2)} \tag{4-10}$$

and the covariance is

$$\phi(\mu,\nu) = \overline{e_\mu e_\nu} = \overline{(\rho e_{\mu-1} + w_\mu) e_\nu} = \rho\phi(\mu-1,\nu) \tag{4-11}$$

The solution to (4-11) is

$$\phi(\mu,\nu) = \phi_{\mu\nu} = \rho^{|\mu-\nu|}\, \overline{e_\mu^2} = \sigma^2 \rho^{|\mu-\nu|} \tag{4-12}$$

For Example 1 we were considering the covariance function

$$\phi(\mu,\nu) = \sigma^2 \exp[-\beta\delta|\mu-\nu|] \tag{4-13}$$

But if the noise is Gaussian, the covariance function completely specifies the random process $\{e_\mu\}$. That is $\{e_\mu\}$ in Example 1 is a Gaussian first order autoregressive scheme with

$$\ln \rho = -.85 \tag{4-14}$$

Now, for this example we may write (using Equation (4-6))

$$w_\mu = r_\mu - \alpha s_\mu - \rho(r_{\mu-1} - \alpha s_{\mu-1}) \; ; \mu = 2,\ldots,N \tag{4-15}$$

But since w_μ are samples of a Gaussian white noise process we can get an approximation to the ML estimates of α and ρ by minimizing, with respect to α and ρ, the quadratic form

$$Q = \sum_{\mu=2}^{N} w_\mu^2 \tag{4-16}$$

(Note: The summation in Q cannot start at $\mu = 1$ because this would require knowing $N + 1$ values of r_μ.)

Proceeding formally by setting $\left.\frac{\partial Q}{\partial \alpha}\right|_{\substack{\alpha=\hat{\alpha} \\ \rho=\hat{\rho}}} = 0$ and $\left.\frac{\partial Q}{\partial \rho}\right|_{\substack{\alpha=\hat{\alpha} \\ \rho=\hat{\rho}}} = 0$

we find that $\hat{\alpha}$ and $\hat{\rho}$ are the solutions to the simultaneous equations

$$\hat{\alpha} = \frac{\sum_{\mu=2}^{N} (r_\mu - \hat{\rho} r_{\mu-1})(g_\mu - \hat{\rho} g_{\mu-1})}{\sum_{\mu=2}^{N} (g_\mu - \hat{\rho} g_{\mu-1})^2} \tag{4-17}$$

$$\hat{\rho} = \frac{\sum_{\mu=2}^{N} (r_\mu - \hat{\alpha}s_\mu)(r_{\mu-1} - \hat{\alpha}s_{\mu-1})}{\sum_{\mu=2}^{N} (r_{\mu-1} - \hat{\alpha}s_{\mu-1})^2} \qquad (4\text{-}18)$$

This system of equations is still difficult to solve simultaneously. However, it is interesting to inspect the structure of $\hat{\alpha}$ and $\hat{\rho}$ in (4-17) and (4-18).

Notice that $\hat{\alpha}$ in (4-17) is very similar to the ML estimate of α in (4-4), and if ρ were known, then the ML estimate of α would be the MVUE of α.

Notice also that the estimate of ρ in (4-18) amounts to subtracting out an estimate of the regression function from the data, and using the known covariance function of the noise (see (4-13) and (4-14)) to construct the estimate. In this particular case, $\hat{\alpha}$ is to be obtained from the simultaneous solution of the above equations and then $\hat{\rho}$ can be calculated. However, the form of the estimate in (4-18) suggests using the LS estimate of α to form $\hat{\rho}$. This then, is a way of motivating the scheme of estimating the covariance parameters by what we will term the "inverse-covariance-function" technique.

4.4 The Inverse-Covariance-Function Technique

This method of estimating the covariance parameters, $\underline{\beta}$, is a

simple extension of the estimate in Equation (4-18). The idea is to obtain an estimate of the noise process by using the LS estimate of $\underline{\alpha}$ to subtract out an estimate of the regression function. Then we treat the estimated noise process as though its covariance function had the functional form given by $\phi(\tau,\underline{\beta})$. In the sampled data case, for example, we can construct estimates

$$\hat{\phi}(p\delta) = \sum_{\mu=p+1}^{N} \hat{e}_\mu \hat{e}_{\mu-p} \; ; \; p = 0,1,\ldots,N-1 \tag{4-19}$$

where

$$\hat{e}_\mu = r_\mu - \hat{\alpha}_{LS} g_\mu \; ; \; \mu = 1,\ldots,N \tag{4-20}$$

(Observe that $\hat{e}_\mu$ is a zero mean Gaussian random variable.)
We then set

$$\hat{\phi}(p\delta) = \phi(p\delta,\underline{\beta})\Big|_{\underline{\beta}=\hat{\underline{\beta}}} \tag{4-21}$$

and solve for $\hat{\underline{\beta}}$ which satisfies (4-21). This requires, of course, that we have as many estimates, $\hat{\phi}$, as we have unknown parameters in $\underline{\beta}$.

Actually, as has been indicated, only parameters in the normalized covariance function are pertinent to improving the LS estimate of $\underline{\alpha}$. Therefore, we will consider estimates of the normalized covariance function, $\theta(\tau,\underline{\beta}) = \phi(\tau,\underline{\beta})/\phi(0,\underline{\beta})$:

$$\hat{\theta}(p\delta) = \frac{\sum_{\mu=p+1}^{N} \hat{e}_\mu \hat{e}_{\mu-p}}{\sum_{\mu=p+1}^{N} \hat{e}^2_{\mu-p}} \quad ; p = 1,\ldots,N-1 \tag{4-22}$$

Henceforth $\underline{\beta}$ will be regarded as a parameter vector of the normalized noise covariance function.

We then set

$$\hat{\theta}(p\delta) = \theta(p\delta,\underline{\beta})\Big|_{\underline{\beta}=\hat{\underline{\beta}}} \tag{4-23}$$

and solve for $\hat{\underline{\beta}}$. This allows us to construct an estimate of $\theta(\tau,\underline{\beta})$, $\hat{\theta}(\tau)$, for the entire range of τ.

To illustrate the procedure we will construct the covariance parameter estimates for Examples 1 and 2.

4.4.1 Estimating ρ in Example 1

In this example we observe that

$$\theta(\delta,\underline{\beta}) = e^{-\beta\delta} = \rho \tag{4-24}$$

and an estimate of ρ will suffice for the purpose of constructing the desired GLS estimator weight matrix. There is only one unknown parameter here, so our estimate is

$$\hat{\rho} = \hat{\theta}(\hat{c}) = \frac{\sum_{\mu=2}^{N} \hat{e}_\mu \hat{e}_{\mu-1}}{\sum_{\mu=2}^{N} \hat{e}^2_{\mu-1}} \tag{4-25}$$

This is identical in form to the estimate in (4-18).

4.4.2 Estimating ω_c in Example 2

Suppose in Example 2 we regard the bandwidth parameter, β, as known, and the center-frequency parameter, ω_c, as the only unknown normalized covariance parameter. Here

$$\theta(\delta,\underline{\beta}) = e^{-\beta\delta} \cos \omega_c\delta \tag{4-26}$$

Again we have only one unknown parameter, and the estimate is

$$\hat{\omega}_c = \frac{1}{\delta} \cos^{-1}\left\{e^{+\beta\delta} \cdot \hat{\theta}(\delta)\right\} \tag{4-27}$$

where $\hat{\theta}(\delta)$ is given by (4-22). (Actually, it is sufficient to estimate the quantity $\omega_c\delta$, since this enables us to construct the desired weight matrix, $\hat{\theta}^{-1}$, in the GLS estimator.)

4.4.3 Estimating Bandwidth and Center-Frequency in Example 2

If we treat ω_c and β as unknown parameters, we need two equations to obtain estimates of these parameters. We observe that

$$\theta(\delta,\underline{\beta}) = e^{-\beta\delta} \cos \omega_c\delta \tag{4-28}$$

$$\theta(2\delta,\underline{\beta}) = e^{-2\beta\delta} \cos 2\omega_c\delta \tag{4-29}$$

This leads to the equation

$$\frac{\hat{\theta}(2\delta)}{\hat{\theta}(\delta)} = \hat{\theta}(\delta)[1-\tan^2(\hat{\omega}_c\delta)] \tag{4-30}$$

Hence,

$$\hat{\omega}_c\delta = \tan^{-1}\left\{[1 - \hat{\theta}(2\delta)/\ \hat{\theta}(\delta)\ ^2]^{1/2}\right\} \tag{4-31}$$

and

$$\hat{\rho} = e^{-\hat{\beta}\delta} = \hat{\theta}(\delta)/\cos(\hat{\omega}_c\delta) \qquad (4\text{-}32)$$

It should be observed that in the above examples there is a possibility that (4-25), (4-31) and (4-32) will lead to illegitimate estimates, for in (4-31) we have no assurance that $\hat{\theta}(2\delta)<[\hat{\theta}(\delta)]^2$, and in (4-25) and (4-32) we cannot be assured that $\hat{\rho}>0$. Furthermore, since these estimates will be computed on a digital computer, it is wise to modify the resulting equations, whenever necessary, to prevent the occurrence of illegal arithmetical machine operations. For example, in (4-31) we can be sure that we take the square root of a non-negative number if we take the absolute magnitude of the quantity $[1 - \hat{\theta}(2\delta)/\{\hat{\theta}(\delta)\}^2]$. While this must be regarded as an artifice, it will prevent a machine abort and provide an answer, which can then be subjected to scrutiny. (Of course, one could provide for an indication by the computer whenever an illegal operation would have resulted if not for the artifice.) It is remarked here that in the simulations of Examples 1 and 2 such artifices were not required.

4.5 The Bias of the Iterative Estimate of $\underline{\alpha}$

Once the parameters of the normalized covariance function are estimated (by any method) we generate the desired weight matrix,

$\hat{\eta} = \hat{\theta}^{-1}$, and the first iteration of the procedure provides us with

$$\underline{\hat{\alpha}}_{I_1} = [G^T\hat{\eta}G]^{-1}G^T\hat{\eta}\underline{R} \tag{4-33}$$

Notice that the mean of $\underline{\hat{\alpha}}_{I_1}$ is

$$\overline{\underline{\hat{\alpha}}_{I_1}} = \underline{\alpha} + \overline{[G^T\hat{\eta}G]^{-1}G^T\hat{\eta}\underline{e}} \tag{4-34}$$

We shall now investigate the bias vector,

$$\underline{b} = \overline{[G^T\hat{\eta}G]^{-1}G^T\hat{\eta}\underline{e}} \tag{4-35}$$

Consider any two nxn matrices, A and B, such that A^{-1} exists. We can express the matrix $[A + B]^{-1}$ as a power series

$$[A+B]^{-1} = A^{-1}[I_n - BA^{-1} + (BA^{-1})^2 - (BA^{-1})^3 + \ldots] \tag{4-36}$$

where I_n denotes the nxn identity matrix. (Equation (4-36) is easily verified by separately pre- and post-multiplying both sides by [A+B].)

Now rewrite (4-35) in the form

$$\underline{b} = \overline{\left[\left(G^T\zeta G\right) + G^T(\hat{\eta} - \zeta)G\right]^{-1}G^T\hat{\eta}\underline{e}} \tag{4-37}$$

where ζ is any NxN matrix for which $[G^T\zeta G]^{-1}$ exists. If we associate $[G^T\zeta G]$ with A in (4-35) and $[G^T(\hat{\eta} - \zeta)G]$ with B, then

$$\underline{b} = [G^T\zeta G]^{-1}E\left\{\left[I_n - \left([G^T(\hat{\eta} - \zeta)G][G^T\zeta G]^{-1}\right) + \left([G^T(\hat{\eta} - \zeta)G][G^T\zeta G]^{-1}\right)^2 - \ldots\right]G^T\hat{\eta}\underline{e}\right\} \tag{4-38}$$

Now suppose that for every pair (μ,ν) where $1\leq\mu\leq N$, $1\leq\nu\leq N$, $\hat{n}_{\mu\nu}$ can be expressed in the form

$$\hat{n}_{\mu\nu} = \hat{k} \sum_{i} \prod_{k_i=1}^{q_i} a_i e_{k_i} \tag{4-39}$$

where q_i is any even integer $\leq N$, the a_i are any real numbers, and $\hat{k}$ is any (real) number (it may be a random variable) which is a common factor of all the elements of $\hat{n}$. This means that every element of $\hat{n}$, normalized to $\hat{k}$, has an expansion into a weighted sum of products involving an even number of factors of zero-mean Gaussian random variables. For instance, consider Example 1. From Equation (3-5) it follows that with $\hat{\rho}$ as in (4-25), $\hat{n}_{\mu\nu}$ satisfies (4-39).

The significance of the property in (4-39) is that when it holds, every term in $\{\ldots\}$ in Equation (4-38) involves a linear combination of products in which there are an odd number of factors of zero-mean Gaussian random variables. Hence when (4-39) holds, $\underline{b} \equiv \underline{0}$, and the first iterated estimate is unbiased for $\underline{\alpha}$. It is easy to see that when (4-39) holds, each successive iteration will also yield an unbiased estimate for $\underline{\alpha}$.

Hence, the iterative estimate for $\underline{\alpha}$ in Example 1 is unbiased.

When (4-39) does not hold it is difficult to establish the bias of the iterative estimate for $\underline{\alpha}$. However, (4-39) merely comprises a

sufficient condition for unbiasedness. It is therefore possible that even if (4-39) is not satisfied the iterative estimate for $\underline{\alpha}$ will be unbiased. Also, the bias may be small, even if it is not exactly zero. In Example 2, for instance, Equation (4-39) is not satisfied when ω_c is treated as an unknown parameter. Nonetheless, as will be seen from the simulation results, the bias of the iterative estimate for $\underline{\alpha}$ is insignificant.

4.6 Simulation Results

We will now move to a discussion of the results of simulating Examples 1 and 2 and employing the iterative estimate of α. We have already described the estimation procedure, and we have given the explicit form of the covariance parameter estimates for each example.

The simulation naturally involves a Monte-Carlo approximation of the desired ensemble expectations. A technique suggested by Levin [43] was used to simulate the desired noise random process. The actual computer programs are presented in Appendix III.

It was decided that since the mean and variance of the LS estimator could be calculated analytically, a comparison of the analytically obtained results with the simulation results for these quantities, would serve as an indication of the reliability of the simulation.

The simulations were carried out using a sampled data approach with ten time samples uniformly spaced on an interval of unit length.

Equation (2-10) was used to calculate the actual variance of the LS estimate of α, and Equation (2-29) was used to calculate the variance of the MVUE. These quantities are denoted by V_{LS} and V_{MIN} respectively. We denote the Monté-Carlo calculation of the mean value of $\hat{\alpha}_{LS}$ by M_{LS} and that for the mean value of the k^{th} iterative estimate of α by M_{IT_k}. The Monté-Carlo approximation for the mean squared error of $\hat{\alpha}_{LS}$ about the true value α_o is represented by MSE_{LS}, while the corresponding quantity for the k^{th} iterative estimate of α is denoted by MSE_{IT_k}. As a measure of the performance of the iterative scheme we use the quantity $PERF_k = V_{MIN}/MSE_{IT_k}$. If the simulation is accurate we should have $0<PERF_k\leq 1$ for all k. In our simulation only two iterations were made. The value of $PERF_2$ is given for these simulations. The calculation of $PERF_2$ was made before truncating the values of V_{MIN} and MSE_{IT_2} to four digits. This accounts for the apparent discrepancy in the data.

The parameters in Example 1 were chosen as follows:*

Case 1. $\alpha_o = 2.0$; $\beta_o = .01$; $k = 100$

Case 2. $\alpha_o = 2.0$; $\beta_o = .01$; $k = 1$

Case 3. $\alpha_o = 2.0$; $\beta_o = .10$; $k = 1$

Case 4. $\alpha_o = 2.0$; $\beta_o = .01$; $\omega_o = 0$

---

*See Equation (2-37) for the definition of k as used here.

The results of 1000 Monté-Carlo runs (requiring just over two minutes of IBM 360 running time for each case above) are presented in Table I.

The simulation for Example 2 is slightly more complicated because the matrix inverse of $\hat{\theta}$ does not have the simple structure it had in Example 1. A computer program for matrix inversion was used to calculate $\hat{\theta}^{-1}$, and two iterations were performed to estimate α. The computer running time was limited to ten minutes for each of the two cases tried, and the maximum number of Monté-Carlo trials which could be obtained within this time period was used. As in Example 1 the simulation employed ten time samples uniformly spaced on an interval of unit length. In case 1 only the center-frequency was regarded as the unknown covariance parameter, and Equation (4-27) was used to estimate it. In case 2, both the bandwidth and center-frequency were estimated using Equations (4-31) and (4-32). In both simulations we used $\alpha_o = 2.0$, $\beta_o = .01$, and $f_c = \omega_c/2\pi$ was set at 1.5 hz corresponding to $k = 1$ in (2-51). The results of the simulation are displayed in Table II.

TABLE I Simulation of Example 1*

CASE	M_{LS}	V_{LS}	MSE_{LS}	V_{MIN}	M_{IT_1}	MSE_{IT_1}	M_{IT_2}	MSE_{IT_2}	$PERF_2$
1	1.9999	.033	.0333	.0001	1.9997	.0001	1.9997	.0001	1.067
2	1.9876	.078	.0788	.0018	2.0016	.0030	2.0001	.0021	.862
3	2.0028	.104	.1021	.0183	1.9984	.0188	1.9990	.0183	.931
4	2.0030	.996	1.0049	.9950	2.0026	1.0063	2.0023	1.0042	.990

TABLE II Simulation of Example 2

CASE	M_{LS}	V_{LS}	MSE_{LS}	V_{MIN}	M_{IT_1}	MSE_{IT_1}	M_{IT_2}	MSE_{IT_2}	$PERF_2$
1#	1.9959	.030	.0263	.0003	1.9994	.0004	1.9994	.0003	.922
2##	1.9937	.030	.0272	.0003	1.9996	.0006	1.9994	.0004	.857

* Note: Except for the quantity $PERF_2$, which is independent of the variance of the noise process, all the mean squared error calculations are normalized to the noise variance.

\# 365 Monté-Carlo runs

\## 387 Monté-Carlo runs

4.7 Discussion of Simulation Results

With the data conveniently summarized in Tables I and II we can make several observations. First we will discuss the data in Table I pertaining to the autoregressive example.

Notice that in all but Case 4 there is a dramatic difference between the variance of the LS estimate and the MVUE; even though these simulations are for the sampled data case. This indicates that the conditions for potential dramatic improvement over the LS estimate given in Equation (2-37), which were derived on the basis of a continuous-time approach, also serve as a guide to such behavior for the sampled data case.* In fact, Case 4 does not satisfy the conditions in (2-37), and indeed, does not offer an example of a case where dramatic improvement over the LS estimate is possible.

As mentioned earlier a comparison between M_{LS} and α_o, and between V_{LS} and MSE_{LS} serves as an indication of the accuracy of the simulation. With this in mind, there is little doubt that the iterative estimator for α significantly improves the LS estimate, in problems

---

*Undoubtedly the extent to which this holds depends upon how "dense" the sampling is. For the cases studied herein the sampling rate was sufficiently high that all of the samples fell within the correlation time of the noise process (the value of τ required for $\phi(\tau,\underline{\beta})/\phi(0,\underline{\beta}) = e^{-1}$).

characterized by Example 1, whenever dramatic improvement is possible. Even in Case 4, where dramatic improvement is not possible, a comparison of MSE_{LS} and MSE_{IT_2} indicates that the iterative estimator is no worse than the LS estimator.

Also, the degree of improvement is such that the variance of the iterative estimate of α is often within a few percent of the variance of the MVUE of α after only two iterations. In Cases 1 and 2 this represents a reduction in the variance of the LS estimate of two orders of magnitude. In Case 3 one order of magnitude reduction in variance is obtained; and this is all that could possibly be obtained in view of the ratio V_{MIN}/V_{LS}.

Coupling the above results with the fact demonstrated earlier that when the inverse-covariance function estimator is used for ρ the iterative estimator of $\underline{\alpha}$ is unbiased, we conclude that the bootstrapping technique for estimating $\underline{\alpha}$, using the inverse-covariance function estimator, is an effective tool for estimating the regression parameter in problems similar to Example 1.

An inspection of Table II, which pertains to the simulation of Example 2, reveals the same outstanding performance. Note that even when the covariance function has two parameters to be estimated, the iterative estimate of $\underline{\alpha}$ achieves two orders of magnitude reduction in

the mean squared estimation error; and again after only two iterations. It should also be observed that the mean value of the iterative estimate of α is approximately equal to α_o. Hence, even though the condition in (4-39) is not satisfied for Example 2, the iterative estimator for α has negligible bias.

It is interesting to observe that in both of the examples simulated, the bootstrapping estimation procedure is capable of reducing the mean squared error of the LS estimator by orders of magnitude even after only one iteration.

4.8 Conclusions

We have demonstrated that in problems involving the estimation of linear regression parameters in colored Gaussian noise, the simple LS estimator can be significantly suboptimal. When the noise covariance function can be described as a known function of a finite number of unknown, nonrandom parameters it is possible to take advantage of this information to improve upon the LS estimator.

By starting with the LS estimator of the regression parameter and employing an iterative bootstrapping procedure, we have shown that it is possible to greatly reduce the mean squared estimation error, even after only one iteration. Furthermore, even though no MVUE for the regression parameter exists unless the normalized noise covariance

function is known precisely, we have seen that the performance of the iterative regression estimate is very near to the performance of the MVUE after only two iterations.

Depending upon the noise covariance, the bootstrapping procedure may lead to an unbiased estimate of the regression parameter; or to an estimate which is approximately unbiased. When the noise is a Gaussian, stationary, first-order autoregressive scheme, the iterative estimator can be rendered unbiased.

The bootstrapping procedure requires an estimate of the pertinent covariance parameters. We have suggested several approaches which may be taken to obtain the desired estimates:

1) maximum likelihood
2) Barankin
3) inverse-covariance function.

Judging from the simulation results, it does not appear that the success of the bootstrapping procedure requires an unbiased estimate of the covariance parameters. To this extent any of the above-mentioned approaches to covariance parameter estimation are admissible, and the Barankin approach, which would provide an unbiased estimate at least for selected points in the covariance parameter space, is possibly superior. In the examples simulated, the inverse-covariance

function technique was employed because of its simplicity.

4.9 Suggestions for Further Study

We have exhibited a class of problems in which the LS estimator of linear regression parameters can be significantly improved upon using a bootstrapping procedure. This class of problems is characterized by the property that the noise spectrum contains a "spike" which has a bandwidth which is small compared to the regression signal's bandwidth, and which is located so that the "mainlobes" of the signal and noise spectra are separated in the frequency domain.

It seems very difficult to state useful necessary and sufficient conditions for the effectiveness of the bootstrapping procedure (or any other procedure, for that matter). It would be desirable, however, to extend the results obtained to other problems where the noise spectrum contains several spikes, for example.

It would also be interesting to find other classes of problems which have the potential of dramatic improvement over the LS estimator, i.e., where the variance of the MVUE is significantly smaller than that of the LS estimator.

Another topic of interest is related to the complicated problem of investigating the sensitivity of the GLS estimator to the choice of the weight matrix [44]. It would be interesting to compare the

performance of the bootstrapping procedure using different covariance parameter estimates.

REFERENCES

1. Rao, C. R. (1967), Linear Statistical Inference and its Applications, Wiley, New York.
2. Grenander, U. and Rosenblatt, M. (1966), Statistical Analysis of Stationary Time Series, Wiley, New York.
3. Parzen, E., "Statistical Inference on Time Series by Hilbert Space Methods, I", Department of Statistics, Stanford University, Technical Report No. 22, January 2, 1959.
4. Parzen, E., "Extraction and Detection Problems and Reproducing Kernel Hilbert Spaces", Journal SIAM Control, Series A, Vol. 1, No. 1, 1962, pp 35-62.
5. Cramér, H. (1963), Mathematical Methods of Statistics, Princeton University Press, Princeton.
6. Wilks, S.S., (1963), Mathematical Statistics, Wiley, New York.
7. Swerling, P., "Parameter Estimation Accuracy Formulas", IEEE Trans. Info. Theory, Vol. IT-10, No. 4, October, 1964, pp 302-314.
8. Swerling, P., "Topics in Generalized Least Squares Signal Estimation", Journal SIAM, Vol. 14, No. 5, Sept., 1966, pp 998-1031.
9. Blackman, R.B. and Tukey, J.W. (1959), The Measurement of Power Spectra from the Point of View of Communication Engineering, Dover Publications, New York.
10. Parzen, E., "On Statistical Spectral Analysis", Proceedings Symposium Applied Math., Vol. 16, 1964, pp 221-246.
11. Levin, M.J., "Power Spectrum Parameter Estimation", IEEE Trans. Info. Theory, Vol. IT-11, Jan., 1965, pp 100-107.

12. Hofstetter, E.M., "Some Results on the Stochastic Signal Parameter Estimation Problem", IEEE Trans. Info. Theory, Vol. IT-11, No. 3, July 1965, pp 422-429.

13. Rao, C.R., "Least Squares Theory Using an Estimated Dispersion Matrix and its Application to Measurement of Signals", Proceedings Fifth Berkeley Symposium on Math. Stat. and Prob., Vol. I, 1967, pp 355-372.

14. Barankin, E.W., "Locally Best Unbiased Estimates", Annals of Math. Stat., Vol. 20, 1949, pp 477-501.

15. Swerling, P., "Parameter Estimation for Waveforms in Additive Gaussian Noise", Journal SIAM, Vol. 7, No. 2, June, 1959, pp 152-166.

16. Gabor, D., Wilby, W. and Woodcock, R., "A Universal Non-Linear Filter, Predictor, and Simulator which Optimizes Itself by a Learning Process", Institution of Electrical Engineers, Proceedings, Part B, Vol. 108, No. 40, July, 1961, pp 422-438.

17. Keehn, D.G., "A Note on Learning for Gaussian Properties", IEEE Trans. Info. Theory, Vol. IT-11, No. 1, Jan., 1965, pp 126-132.

18. Widrow, B., "Adaptive Filters I: Fundamentals", Stanford Electronics Lab., Tech. Report No. 6764-6, Dec., 1966.

19. Dressler, R.M., "A Simplified Technique for the Synthesis of Model-Referenced Adaptive Control Systems", Stanford Electronics Lab., Tech. Report No. 6764-3, March, 1966.

20. Ula, N. and Kim, M., "An Empirical Bayes Approach to Adaptive Control", Journal Franklin Institute, Vol. 280, No. 3, Sept., 1965, pp 189-204.

21. Widrow, B., Mantey, P.E., et.al., "Adaptive Antenna Systems", Proc. IEEE, Vol. 55, No. 12, Dec., 1967, pp 2143-2159.

22. Conference on Adaptive Telemetry, held Feb. 15-16, 1966, Goddard Space Flight Center, Jan., 1967.

23. Huggett, W., Gronquist, R., et.al., "Optimal Processing of Antenna Arrays for Simultaneously Resolving Many Signals in Both Bearing and Frequency", Univ. Michigan, Nov., 1964.

24. Albert, A. and Sittler, R., "A Method for Computing Least-Squares Estimators that Keep Up with the Data", Journal SIAM Control, Series A, Vol. 3, No. 3, 1966, pp 384-417.

25. Jazwinski, A. and Bailie, A., "Adaptive Filtering", Analytical Mechanics Associates, Inc., Report No. 67-6, March, 1967.

26. Khaitman, E.N., "Optimal Reception in the Absence of A-priori Information on the Statistical Noise Properties", Telecommunications and Radio Engineering, Vol. 20, No. 9, Sept., 1966, pp 40-45.

27. Carlton, A.G., and Follin, J.W., Jr., "Recent Developments in Fixed and Adaptive Filtering", Proceedings Second AGARD Guided Missiles Seminar (Guidance and Control), AGAR Dograph 21, Sept., 1956.

28. Cox, H., "On the Estimation of State Variables and Parameters for Noise Dynamic Systems", IEEE Trans. Auto. Control, PGAC Vol. AC-9, No. 1, Jan., 1964, pp 5-12.

29. Levine, N., "A Technique for Increasing the Flexibility of Recursive Least Squares Data Smoothing", Bell Sys. T.J., Vol. 40, May, 1961, pp 821-840.

30. Marquardt, D.W., "An Algorithm for Least-Squares Estimation of Non-Linear Parameters", Journal SIAM, June, 1963, pp 431-441.

31. Sakrison, D.J., "Iterative Design of Optimum Filters for Non-Mean-Square Error Performance Criterion", IEEE Trans. Info. Theory, PGIT Vol. IT-9, No. 3, July, 1963, pp 161-167.

32. Goldmann, J.B., Adaptive Control Systems: An Annotated Bibliography, Lockheed, Calif., AD-433807, Sept., 1963.

33. Kennedy, R. and O'Neill, R., "Adaptive Beamforming for Sonar Signal Processing", U.S. Navy Underwater Sound Lab., Tech. Memorandum No. 914-184-66, June 29, 1966.

34. Eicker, F., "Limit Theorems for Regressions with Unequal and Dependent Errors", Proc. Fifth Berkeley Symposium on Math. Stat. and Prob., Vol. I, 1967, pp 59-82.

35. Swerling, P., "Optimum Linear Estimation for Random Processes as the Limit of Estimates Based on Sampled Data", IRE Wescon Conv. Record, Part 4, 1958, pp 158-163.

36. Grenander, U. and Szego, G. (1958), Toeplitz Forms and Their Applications, UCLA Press, Berkeley.

37. Zemanian, A.H., (1965), Distribution Theory and Transform Analysis, McGraw-Hill, New York.

38. Aĭvazjan, S.A. and Rozanov, Ju, A., "Some Remarks on Assymptotically Effective Linear Estimates of Regression Coefficients", Selected Translations in Math. Stat. and Prob., Vol. 6, 1966, pp 37-50.

39. Lehmann, E.L. and Scheffe, H., "Completeness, Similar Regions, and Unbiased Estimation—Part 1", Sankhya, Vol. 10, Part 4, 1950, pp 305-340.

40. Lehmann, E.L. and Scheffe, H., "Completeness, Similar Regions, and Unbiased Estimation—Part 2", Sankhya, Vol. 15, Part 3, 1955, pp 219-236.

41. Siddiqui, M.M., "On the Inversion of the Sample Covariance Matrix in a Stationary Autoregressive Process", Annals of Math. Stat., Vol. 129, June, 1958, pp 585-588.

42. Bartlett, M.S., (1966), Stochastic Processes, Cambridge University Press, Cambridge.

43. Levin, M.J., "Generation of a Sampled Gaussian Time Series Having a Specified Correlation Function", IRE Trans. Info. Theory, Vol. IT-6, No. 5, Dec., 1960, pp 545-548.

44. Nash, R. and Tuteur, F., "The Effect of Uncertainties in the Noise Covariance Matrices on the Maximum Likelihood Estimate of a Vector", IEEE Trans. Auto. Control, Vol. AC-13, No. 1, Feb., 1968, pp 86-88.

45. Zadeh, L.A. and Raggazzini, J.R., "An Extension of Wiener's Theory of Prediction", Journal Appl. Phys., Vol. 21, July, 1950, pp 645-655.

46. Zadeh, L.A. and Raggazzini, J.R., "Optimum Filters for the Detection of Signals in Noise", Proc. IRE, Vol. 40, Oct., 1952, pp 1223-1231.

47. Swerling, P., "The Greatest Lower Bound for the Variance of Unbiased Estimates", Report R-90, Control Systems Lab., Univ. Illinois, April, 1957.

48. Hewitt, E. and Stromberg, K., (1965), Real and Abstract Analysis, Springer-Verlag, New York.

49. Kelly, E., Reed, I., and Root, W., "The Detection of Radar Echoes in Noise, I", Journal SIAM, Vol. 8, No. 2, June, 1960, pp 309-341.

50. Kelly, E., Reed, I., and Root, W., "The Detection of Radar Echoes in Noise, II", Journal SIAM, Vol. 8, No. 3, September, 1960, pp 481-507.

51. Doob, J.L., (1953), Stochastic Processes, Wiley, New York.

52. Bartlett, M.S., (1966), An Introduction to Stochastic Processes, Cambridge University Press, Cambridge.

53. Van Trees, H.L., (1968), Detection, Estimation, and Modulation Theory, Part I, Wiley, New York.

54. Harger, R.O., "Maximum Likelihood Estimation of Spectral Density Parameters of Gaussian Processes", Radar and Optics Lab., Willow Run Lab., Institute of Science and Technology, Univ. of Michigan, Ann Arbor, Michigan, Nov., 1966.

55. Brennan, L.E. and Reed, I.S., "Optimum Processing of Unequally Spaced Radar Pulse Trains for Clutter Rejection", IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-4, No. 3, May, 1968, pp 474-477.

APPENDICES

APPENDIX I

In Section 2.4 we had need for the solution, h(t), of the integral equation

$$\sigma^2 \int_0^T h(t) \exp(-\beta|t-s|) \cos[\omega_c(t-s)]\, dt = 1 \; ; \; s\varepsilon[0,T] \qquad (I-1)$$

Here we present the details for finding h(t).

According to Zadeh and Raggazzini, [45, 46], by virtue of Equation (2-62) the solution is of the form:

$$h(t) = K_o e^{-kt} + K_1 + K_2\delta(t) + K_3\delta(t-T) + K_4 e^{-k(T-t)} \qquad (I-2)$$

We will find the six constants by direct substitution of (I-2) into (I-1). To facilitate the algebra we will perform the integration for each term of h(t) separately.

Hence, integrating with the kernel $K_o e^{-kt}$ we obtain,

$$I_1 = \sigma^2 \int_0^T K_o e^{-kt}\, e^{-\beta|t-s|} \cos[\omega_c(t-s)]\, dt$$

$$= \sigma^2 K_o \int_0^s e^{-kt}\, e^{-\beta(s-t)} \cos[\omega_c(s-t)]\, dt$$

$$+ \sigma^2 K_o \int_s^T e^{-kt}\, e^{+\beta(s-t)} \cos[\omega_c(s-t)]\, dt$$

$$= I_{11} + I_{12} \qquad (I-3)$$

where

$$I_{11} = \frac{K_o\sigma^2(k-\beta)e^{-ks}}{\omega_c^2 + (k-\beta)^2}\left\{-1 + e^{-(k-\beta)s}\left[\cos\omega_c s + \frac{\omega_c}{k-\beta}\sin\omega_c s\right]\right\} \qquad (I\text{-}4)$$

$$I_{12} = \frac{K_o\sigma^2(k+\beta)e^{-ks}}{\omega_c^2 + (k+\beta)^2}\left\{1 + e^{-(k+\beta)(T-s)}\left[-\cos[\omega_c(T-s)] + \frac{\omega_c}{k+\beta}\sin[\omega_c(T-s)]\right]\right\} \qquad (I\text{-}5)$$

The integration with K_1 as the kernel may be obtained from the above by setting $k = 0$. Therefore;

$$I_2 = \sigma^2 \int_0^T K_1\, e^{-\beta|t-s|}\cos[\omega_c(t-s)]\, dt = I_{21} + I_{22} \qquad (I\text{-}6)$$

where,

$$I_{21} = \frac{\sigma^2 K_1\beta}{\omega_c^2 + \beta^2}\left(1 - e^{-\beta s}\left[\cos\omega_c s - \frac{\omega_c}{\beta}\sin\omega_c s\right]\right) \qquad (I\text{-}7)$$

$$I_{22} = \frac{\sigma^2 K_1\beta}{\omega_c^2 + \beta^2}\left(1 + e^{-\beta(T-s)}\left[-\cos[\omega_c(T-s)] + \frac{\omega_c}{\beta}\sin[\omega_c(T-s)]\right]\right) \qquad (I\text{-}8)$$

The integrals resulting from the third and fourth terms in (I-2) are easily seen to be

$$I_3 = K_2\sigma^2 e^{-\beta s}\cos\omega_c s \qquad (I\text{-}9)$$

$$I_4 = K_3\sigma^2 e^{-\beta(T-s)}\cos[\omega_c(T-s)] \qquad (I\text{-}10)$$

Finally, the result of using $K_4 e^{-k(T-t)}$ as the kernel may be found from (I-4) and (I-5) upon replacing k by -k and K_o by $K_4 e^{-kT}$. That is,

$$I_5 = \sigma^2 \int_0^T K_4 e^{-k(T-t)}\ e^{-\beta|t-s|}\ \cos[\omega_c(t-s)]\ dt = I_{51} + I_{52} \qquad \text{(I-11)}$$

where,

$$I_{51} = -\frac{K_4\sigma^2(k+\beta)e^{+ks}e^{-kT}}{\omega_c^2 + (k+\beta)^2}\left(-1 + e^{-(k+\beta)s}\left[\cos\omega_c s - \frac{\omega_c}{k+\beta}\sin\omega_c s\right]\right) \qquad \text{(I-12)}$$

$$I_{52} = -\frac{K_4\sigma^2(k-\beta)e^{+ks}e^{-kT}}{\omega_c^2 + (k-\beta)^2}\left(1 + e^{-(\beta-k)(T-s)}\left[-\cos[\omega_c(T-s)] - \frac{\omega_c}{k-\beta}\sin[\omega_c(T-s)]\right]\right) \qquad \text{(I-13)}$$

We now solve for the constants by requiring that

$$I_1 + I_2 + I_3 + I_4 + I_5 = 1 \qquad \text{(I-14)}$$

The right-hand side of (I-14) has no dependence on s. Hence, the constants on the left-hand side must be chosen so that the dependence on s vanishes. For example, it can be verified that the terms in $e^{\pm ks}$ which appear on the left-hand side of (I-14) will vanish for arbitrary K_o and K_4 (different from zero) if

$$k^2 = \omega_c^2 + \beta^2 \qquad \text{(I-15)}$$

To solve for K_o and K_4 we equate the coefficients of $e^{-\beta s}\sin\omega_c s$ and $e^{-\beta(T-s)}\sin[\omega_c(T-s)]$ to zero. This results in the equations

$$\frac{K_o\omega_c}{\omega_c^2 + (k-\beta)^2} + \frac{K_1\omega_c}{\omega_c^2 + \beta^2} + \frac{K_4\omega_c e^{-kT}}{\omega_c^2 + (k+\beta)^2} = 0 \tag{I-16}$$

$$\frac{K_o\omega_c e^{-kT}}{\omega_c^2 + (k+\beta)^2} + \frac{K_1\omega_c}{\omega_c^2 + \beta^2} + \frac{K_4\omega_c}{\omega_c^2 + (k-\beta)^2} = 0 \tag{I-17}$$

This system of equations yields

$$K_o = K_4 \tag{I-18}$$

and

$$K_o\left[\frac{1}{\omega_c^2 + (k-\beta)^2} + \frac{e^{-kT}}{\omega_c^2 + (k+\beta)^2}\right] + K_1\frac{1}{\omega_c^2 + \beta^2} = 0 \tag{I-19}$$

But from the constant terms on both sides of (I-14) we have

$$K_1 = \frac{\omega_c^2 + \beta^2}{2\beta\sigma^2} \tag{I-20}$$

Inserting this result into (I-19) yields

$$K_o = -\frac{[\omega_c^2 + (k-\beta)^2][\omega_c^2 + (k+\beta)^2]}{2\beta\sigma^2\left([\omega_c^2 + (k+\beta)^2] + [\omega_c^2 + (k-\beta)^2]e^{-kT}\right)} \tag{I-21}$$

Finally, setting the coefficients of $e^{-\beta s}\cos\omega_c s$ and $e^{-\beta(T-s)}\cos\omega_c(T-s)$ equal to zero provides us with exactly the same

equations for K_2 and K_3; so that

$$K_2 = K_3 \tag{I-22}$$

and

$$\frac{K_o\sigma^2(k-\beta)}{\omega_c^2 + (k-\beta)^2} - \frac{1}{2} + K_2\sigma^2 - \frac{K_o\sigma^2(k+\beta)e^{-kT}}{\omega_c^2 + (k+\beta)^2} = 0 \tag{I-23}$$

Hence,

$$K_2 = \frac{k\left([\omega_c^2 + (k+\beta)^2] - [\omega_c^2 + (k-\beta)^2]e^{-kT}\right)}{2\beta\sigma^2\left([\omega_c^2 + (k+\beta)^2] + [\omega_c^2 + (k-\beta)^2]e^{-kT}\right)} \tag{I-24}$$

Using (I-15) these results may be expressed in the form of Equations (2-55) through (2-59).

APPENDIX II

In Chapter III we made use of Barankin's theory of locally best unbiased estimation. Here we give details of the development of this subject; following closely the approaches of Barankin [14] and Swerling [15, 47].

We begin by defining a measure space $(X,\mathcal{F},\mu)$ where X is a sample space of points, x, $\mathcal{F}$ is a σ-field of subsets of X, and μ is a countably additive measure defined on $\mathcal{F}$.

Consider the family of probability measures $\{P_\alpha;\alpha\varepsilon A\}$ defined on $\mathcal{F}$ where α is a real scalar parameter belonging to the set A. We shall assume that the measures P_α are absolutely continuous with respect to the measure μ for every $\alpha\varepsilon A$. Hence, we may define the family of probability density functions, $\mathcal{P} = \{p_\alpha;\alpha\varepsilon A\}$, on X with respect to the measure μ using the Radón-Nikodym theorem. So, let B be any subset of X. Then

$$\int_B dP_\alpha = \int_B p_\alpha d\mu \quad \forall\ B\subset X \text{ and } \alpha\varepsilon A \tag{II-1}$$

In our application we shall take μ to be Lebesgue measure, and $\mathcal{P}$ will be the family of Gaussian probability densities defined on the extended real line. We shall retain the above notation, however, for the sake of generality.

Our goal is to draw inferences on the unknown nonrandom parameter,

α, or on some known (real valued) function g of this parameter. That is, let g be a known real valued function defined on A. We seek a real valued μ-measurable function, $T_o(x)$, on X which is an unbiased estimate of $g(\alpha)$, and which is best in the mean square sense at some prescribed point $\alpha = \alpha_o$. Thus, if we let $\mathcal{U}$ denote the class of all μ-measurable functions, $T(x)$, on X having the unbiasedness property

$$\int_X T dP_\alpha = g(\alpha) \qquad \forall \alpha \in A \tag{II-2}$$

we seek $T_o \in \mathcal{U}$ such that

$$\int_X [T_o - g(\alpha_o)]^2 \, dP_{\alpha_o} \leq \int_X [T - g(\alpha_o)]^2 \, dP_{\alpha_o} \qquad \forall T \in \mathcal{U} \tag{II-3}$$

It is convenient to define

$$\phi(x) = T(x) - g(\alpha_o) \tag{II-4}$$

$$h(\alpha) = g(\alpha) - g(\alpha_o) \tag{II-5}$$

We also define the norm on X with respect to the nominal measure P_{α_o} for any μ-measurable function, ψ as

$$\| \psi \| = \left[\int_X \psi^2(x) dP_{\alpha_o}(x) \right]^{1/2} \tag{II-6}$$

Note that $\| \phi \|^2$ is the mean squared error of the estimate T about $g(\alpha_o)$.

Now for $T \in \mathcal{U}$ we can rewrite (II-2) as

$$\int_X \phi(x)dP_\alpha(x) = h(\alpha) \qquad \forall \alpha \in A \tag{II-7}$$

That is, $T(x)$ is an unbiased estimator for $g(\alpha)$ if and only if $\phi(x)$ is an unbiased estimator for $h(\alpha)$.

We shall now assume that each of the ratios

$$\pi(\alpha,x) = \frac{p_\alpha(x)}{p_{\alpha_o}(x)} \tag{II-8}$$

is defined μ-almost everywhere on X, and that $\pi(\alpha,x)$, considered as a function of α, is Borel measurable. Then Equation (II-7) takes the form

$$\int_X \phi(x)\pi(\alpha,x)dP_{\alpha_o}(x) = h(\alpha) \; ; \; \forall \alpha \in A \tag{II-9}$$

Let us now consider the measurable space $(A,\mathcal{A})$, and let λ be any signed measure on $(A,\mathcal{A})$ such that the integrals $\int_A h(\alpha)d\lambda(\alpha)$ and $\iint_{X\otimes A} \pi(\alpha,x)d\lambda(\alpha)dP_{\alpha_o}(x)$ exist. Then we may integrate both sides of (II-9) with respect to λ and obtain

$$\left|\int_A h(\alpha)d\lambda(\alpha)\right| = \left|\int_A \int_X \pi(\alpha,x)\phi(x)d\lambda(\alpha)dP_{\alpha_o}(x)\right|$$

$$\leq \|\phi\| \cdot \left\|\int_A \pi(\alpha,x)d\lambda(\alpha)\right\| \tag{II-10}$$

where the inequality follows from an application of the Schwarz inequality.

This leads to the result that the greatest lower bound on the quantity $\|\phi\|^2$ is:

$$\|\phi\|^2_{glb} = \underset{\lambda}{lub} \left\{ \frac{\left[\int_A h(\alpha) d\lambda(\alpha)\right]^2}{\left\|\int_A \pi(\alpha,x) d\lambda(\alpha)\right\|^2} \right\} \qquad \text{(II-11)}$$

The least upper bound of the expression on the right-hand side of (II-11) must be taken over all admissible measures λ in order to obtain $\|\phi\|^2_{glb}$. However, the insertion of <u>any</u> admissible λ on the right-hand side will furnish <u>a</u> lower bound for $\|\phi\|^2$, though not necessarily the greatest lower bound.

The quantity $\|\phi\|^2_{glb}$ is referred to as the Barankin bound. When α_o is regarded as the true value of the parameter α, Equation (II-11) provides us with the greatest lower bound on the variance of all unbiased estimates of $g(\alpha)$.

Our main concern here is with the problem of evaluating the Barankin bound. We are also interested in obtaining an explicit construction for the estimator which attains this bound (hereafter referred to as the Barankin estimator).

To this end, for any given α_o let $\mathcal{U}_2^{(\alpha_o)}$ denote the set of all functions, $\phi(x)$ satisfying (A-7) for which $\|\phi\|^2 < \infty$. Assume that $\mathcal{U}_2^{(\alpha_o)}$ is nonempty. Let $\overline{\mathcal{U}_2^{(\alpha_o)}}$ denote the closure of $\mathcal{U}_2^{(\alpha_o)}$ taken with $\| \ \|^2$ as the metric. Since $\overline{\mathcal{U}_2^{(\alpha_o)}}$ is closed and convex, there exists a function, $\phi_o(x)$, (and corresponding to it, $T_o(x)$) which minimizes $\|\phi\|^2$ for all functions in $\overline{\mathcal{U}_2^{(\alpha_o)}}$ and which is unique with probability one (P_α measure for any $\alpha \varepsilon A$). It is remarked that in general $T_o(x)$ will depend upon α_o. This will be the case unless $T_o(x)$ minimizes $\|\phi\|^2$ for every $\alpha_o \varepsilon A$.

It will be assumed that $T_o(x)$ is an unbiased estimate of $g(\alpha)$, for if this were not the case, there would be no unbiased estimate of $g(\alpha)$ which had minimum mean squared error when $\alpha = \alpha_o$.

Consider now functions $f(\alpha)$ defined over A having the form

$$f(\alpha) = \int_X H(x) dP_\alpha(x) \qquad \forall \alpha \varepsilon A \tag{II-12}$$

where $H(x)$ is any μ-measurable function such that $\|H\|^2 < \infty$. Let F denote the family of all such functions. Then, it is easy to show that for any $f \varepsilon F$ and for <u>any</u> concomitant H, the integral $\int_X T_o(x) H(x) dP_{\alpha_o}(x)$ yields the <u>same</u> constant. To see this, for some

(real) ε define the μ-measurable function

$$T^*(x) = T_o(x) + \varepsilon[H_1(x) - H_2(x)] \qquad \text{(II-13)}$$

where H_1 and H_2 both satisfy (II-12) with $\|H_1\|^2 < \infty$ and $\|H_2\|^2 < \infty$. Since H_1 and H_2 satisfy (II-12), $T^*(x)$ is in $\mathcal{U}$. The fact that $\|H_1\|^2$ and $\|H_2\|^2$ are bounded ensures that $T^*(x)$ is in $\overline{\mathcal{U}_2^{(\alpha_o)}}$. Consider

$$D(\varepsilon) = \int_X [T^*(x) - g(\alpha_o)]^2 \, dP_{\alpha_o}(x) \qquad \text{(II-14)}$$

Calculating $\frac{\partial D(\varepsilon)}{\partial \varepsilon}$ will reveal that unless

$$\int_X T_o(x)[H_1(x) - H_2(x)] \, dP_{\alpha_o}(x) = 0 \qquad \text{(II-15)}$$

we will have $\frac{\partial D(\varepsilon)}{\partial \varepsilon}\Big|_{\varepsilon=0} \neq 0$. This, however, would contradict the definition of $T_o(x)$.

We may now define a functional $\Lambda^{(\alpha_o)}$ on F as follows. For

$$f(\alpha) = \int_X H(x) dP_\alpha(x) \; ; \; \alpha \varepsilon \Lambda, \; \|H\|^2 < \infty$$

define

$$\Lambda^{(\alpha_o)}[f] = \int_X H(x)\left[T_o(x) - g(\alpha_o)\right] dP_{\alpha_o}(x) \qquad \text{(II-16)}$$

The fact that $\int_X T_o(x)H(x)dP_{\alpha_o}(x)$ is a constant ensures that $\Lambda^{(\alpha_o)}[f]$ is uniquely defined for any $f \varepsilon F$ and for any concomitant H. Note that

if f_1 and f_2 are elements of F, $\Lambda^{(\alpha_o)}[f_1 + f_2] = \Lambda^{(\alpha_o)}[f_1] + \Lambda^{(\alpha_o)}[f_2]$. Also for any constant k and for all $f \in F$ we have $\Lambda^{(\alpha_o)}[kf] = k\Lambda^{(\alpha_o)}[f]$. Therefore, $\Lambda^{(\alpha_o)}$ is a linear functional defined on F.

Two observations are now in order. First we notice that for $f(\alpha) = c$, where c is any (real) constant, we may use $H(x) = c$ in (II-12). Hence, for any constant, c

$$\Lambda^{(\alpha_o)}[c] \equiv 0 \tag{II-17}$$

Also, setting $f(\alpha) = g(\alpha)$ we may use $H(x) = T_o(x)$ and discover that

$$\Lambda^{(\alpha_o)}[g] = \|\phi\|^2_{glb} = \|\phi_o\|^2 \tag{II-18}$$

This shows that evaluation of the Barankin bound is equivalent to evaluating $\Lambda^{(\alpha_o)}[g]$. We will now show that it is possible to evaluate $\Lambda^{(\alpha_o)}$ on a certain set of functions of α to be defined below. Once this is done, if $g(\alpha)$ can be expressed as a linear combination of the functions for which $\Lambda^{(\alpha_o)}[\cdot]$ can be evaluated, we will be able to calculate $\Lambda^{(\alpha_o)}[g]$.

Define the function

$$G(\alpha,\alpha'|\alpha_o) = \int_X \pi(\alpha,x)\pi(\alpha',x)dP_{\alpha_o}(x) \tag{II-19}$$

and assume that $G(\alpha,\alpha'|\alpha_o)$ is bounded for all $\alpha\varepsilon A$ whenever α' belongs to some set $A_{\alpha_o} \subset A$. It is easy to see that $G(\alpha,\alpha'|\alpha_o)$, considered as a function of α can be expressed in the form

$$G(\alpha,\alpha'|\alpha_o) = \int_X H(x)dP_\alpha(x) \; ; \quad \|H\| < \infty$$

if $H(x)$ is identified as

$$H(x) = \pi(\alpha',x) \tag{II-20}$$

and if $\|H\|^2 < \infty$ whenever $\alpha'\varepsilon A_{\alpha_o}$.

Hence, for $\Lambda^{(\alpha_o)}$ operating on G considered as a function of α, we have

$$\Lambda^{(\alpha_o)}[G(\alpha,\alpha'|\alpha_o)] = \int_X \pi(\alpha',x)\left[T_o(x) - g(\alpha_o)\right] dP_{\alpha_o}(x)$$

$$= g(\alpha') - g(\alpha_o) \; ; \; \alpha'\varepsilon A_{\alpha_o} \tag{II-21}$$

Now suppose for the moment that A is composed of discrete points $\{\alpha_i \; ; \; i = 0,1,\ldots,n\}$ (n may be infinite). Also suppose that $g(\alpha)$ on A can be expressed as a linear combination of the functions $G(\alpha,\alpha'|\alpha_o)$ considered as functions of α' for $\alpha'\varepsilon A_{\alpha_o}$. That is, suppose there exists some set of real numbers $\{a_i\}$ such that for $\{\alpha_i\}\varepsilon A$

$$g(\alpha_i) - g(\alpha_o) = \sum_{j=o}^{n} a_j G(\alpha_i,\alpha_j|\alpha_o) \; ; \; \{\alpha_j\}\varepsilon A_{\alpha_o} \tag{II-22}$$

(The set $\{a_i\}$ may be found by inverting the $(n + 1) \times (n + 1)$ matrix of elements $[G(\alpha_i,\alpha_j|\alpha_o)\ ;\ i,j = 0,1,\ldots,n]$). Then

$$\|\phi_o\|^2 = \Lambda^{(\alpha_o)}[g] = \Lambda^{(\alpha_o)}[g - g(\alpha_o)] = \sum_{j=o}^{n} a_j \Lambda^{(\alpha_o)}\left[G(\alpha_i,\alpha_j|\alpha_o)\right]$$

$$= \sum_{j=o}^{n} a_j \left[g(\alpha_j) - g(\alpha_o)\right] \quad \text{(II-23)}$$

Equation (II-23) provides us with an exact expression for the Barankin bound when the parameter space is composed of a denumerable set of points.

We can verify that the estimator

$$\phi(x) = \sum_{j=o}^{n} a_j \pi(\alpha_j,x)\ ;\ \{\alpha_j\} \varepsilon A_{\alpha_o} \quad \text{(II-24)}$$

is unbiased for $h(\alpha)$ when the $\{a_i\}$ satisfy (II-22), since for $\alpha_i \varepsilon A$

$$\int_X \phi(x) dP_{\alpha_i}(x) = \sum_{j=o}^{n} a_j \int_X \frac{p(x;\alpha_i)p(x;\alpha_j)}{p(x;\alpha_o)}\, dx$$

$$= \sum_{j=o}^{n} a_j G(\alpha_i,\alpha_j|\alpha_o) = g(\alpha_i) - g(\alpha_o)\ ;\ \{\alpha_i\} \varepsilon A \quad \text{(II-25)}$$

Also,

$$\|\phi(x)\|^2 = \int_X \sum_{j=o}^{n} \sum_{k=o}^{n} a_j a_k \pi(\alpha_j,x)\pi(\alpha_k,x)p(x;\alpha_o)dx$$

$$= \sum_{j=o}^{n} a_j \sum_{k=o}^{n} a_k G(\alpha_j,\alpha_k|\alpha_o) = \sum_{j=o}^{n} a_j \Big[g(\alpha_j) - g(\alpha_o) \Big] \equiv \|\phi_o\|^2 \qquad \text{(II-26)}$$

Hence, by the uniqueness property of the Barankin estimator

$$T_o(x) = \phi_o(x) + g(\alpha_o) = \sum_{j=o}^{n} a_j \pi(\alpha_j,x) + g(\alpha_o) \qquad \text{(II-27)}$$

So far we have shown how to calculate the Barankin bound and the Barankin estimator when A is composed of a countable set of points. We now want to generalize these results to the case when A is an interval on the real line. Loosely speaking, what is needed is a representation for $g(\alpha)$ on A, similar to the expression in (II-22); so that the linear functional $\Lambda^{(\alpha_o)}$ operating on $g(\alpha)$ can be expressed in terms of $\Lambda^{(\alpha_o)}$ operating on $G(\alpha,\alpha'|\alpha_o)$ considered as a function of α. First we present some preliminaries.

Suppose that $f_i(\alpha)$ is any function defined on A which belongs to F. Suppose that for every i there is some $H_i(x)$ with $\|H_i\|^2 < \infty$ such that with

$$f_i(\alpha) = \int_X H_i(x) dP_\alpha(x) \; ; \; \alpha \epsilon A \qquad \text{(II-28)}$$

the quantities

$$f_\infty(\alpha) = \lim_{i\to\infty} f_i(\alpha) \tag{II-29}$$

$$H_\infty(x) = \lim_{i\to\infty} H_i(x) \tag{II-30}$$

exist and $\lim_{i\to\infty} \|H_i - H_\infty\|^2 = 0$ (That is, H_i converges in mean to H_∞).

Then,

$$0 \leq \left|\Lambda^{(\alpha_o)}\left[f_\infty\right] - \lim_{i\to\infty} \Lambda^{(\alpha_o)}\left[f_i\right]\right| = \left|\lim_{i\to\infty} \int_X \left[T_o(x) - g(\alpha_o)\right] \times \left[H_\infty(x) - H_i(x)\right] dP_{\alpha_o}(x)\right| \leq \|\phi_o\|^2 \cdot \lim_{i\to\infty} \|H_i - H_\infty\|^2 = 0 \tag{II-31}$$

Hence, under the conditions stated above

$$\Lambda^{(\alpha_o)}\left[\lim_{i\to\infty} f_i\right] = \lim_{i\to\infty} \Lambda^{(\alpha_o)}\left[f_i\right] \tag{II-32}$$

Consider now any function or generalized function, $d\lambda$, on A_{α_o} such that both integrals $\int_X \pi(\alpha,x) \int_{A_{\alpha_o}} \pi(\alpha',x) d\lambda(\alpha') dP_{\alpha_o}(x)$, $\int_{A_{\alpha_o}} G(\alpha,\alpha'|\alpha_o) d\lambda(\alpha')$ exist and are equal for all $\alpha \varepsilon A$. Suppose, corresponding to (II-22), we can find such a $d\lambda$, say $d\overset{o}{\lambda}$, for which

$$\pi(\alpha) - g(\alpha_o) = \int_{A_{\alpha_o}} G(\alpha,\alpha'|\alpha_o) d\overset{o}{\lambda}(\alpha') \; ; \; \forall \alpha \varepsilon A \tag{II-33}$$

That is, $g(\alpha) - g(\alpha_o)$ has the representation

$$g(\alpha) - g(\alpha_o) = \int_X \left\{ \int_{A_{\alpha_o}} \pi(\alpha',x) d\mathring{\lambda}(\alpha') \right\} dP_\alpha(x) \qquad \text{(II-34)}$$

Identifying $\int_{A_{\alpha_o}} \pi(\alpha',x) d\mathring{\lambda}(\alpha')$ with $H(x)$ in (II-12), we see that if

$\left\| \int_{A_{\alpha_o}} \pi(\alpha',x) d\mathring{\lambda}(\alpha') \right\|^2 < \infty$, then $[g(\alpha) - g(\alpha_o)] \varepsilon F$, and $\Lambda^{(\alpha_o)}[g]$ is

uniquely defined. Therefore,

$$\begin{aligned} \|\phi\|^2_{glb} &= \Lambda^{(\alpha_o)}[g] = \Lambda^{(\alpha_o)}[g - g(\alpha_o)] \\ &= \int_X \left\{ \int_{A_{\alpha_o}} \pi(\alpha',x) d\mathring{\lambda}(\alpha') \right\} \cdot \left[T_o(x) - g(\alpha_o) \right] dP_{\alpha_o}(x) \\ &= \int_{A_{\alpha_o}} \int_X \left[T_o(x) - g(\alpha_o) \right] dP_{\alpha'}(x) d\mathring{\lambda}(\alpha') \\ &= \int_{A_{\alpha_o}} \left[g(\alpha') - g(\alpha_o) \right] d\mathring{\lambda}(\alpha') \qquad \text{(II-35)} \end{aligned}$$

The Barankin estimator in this case becomes

$$T_o(x) = \int_{A_{\alpha_o}} \pi(\alpha,x) d\mathring{\lambda}(\alpha) + g(\alpha_o) \qquad \text{(II-36)}$$

Furthermore, in view of (II-32), if we have an infinite sequence $\{g_i(\alpha)\}$ converging to $g(\alpha)$ $\forall \alpha \in A$ and a concomitant sequence $\{d\lambda_i^o\}$ satisfying (II-33) and such that for some $d\lambda^o$, $\lim_{i\to\infty} \left\| \int_{A_{\alpha_o}} \pi(\alpha,x) d\lambda_i^o(\alpha) - \int_{A_{\alpha_o}} \pi(\alpha,x) d\lambda(\alpha) \right\|^2 = 0$, then (II-35) and (II-36) remain valid.

While it is admitted that in many cases we cannot hope to find a closed form solution for $d\lambda^o$ satisfying (II-33), two facts are of interest.

First, as has been mentioned, the use of <u>any</u> admissible λ will furnish <u>a</u> lower bound for $\|\phi\|^2$ via (II-11) without the lub operation. In fact, to this end we need not restrict λ to be a signed measure. That is, if $d\lambda$ is any function or generalized function for which

$$\int_X \int_A \pi(\alpha,x) d\lambda(\alpha) \phi(x) dP_{\alpha_o}(x) = \int_A \int_X \phi(x) \pi(\alpha,x) dP_{\alpha_o}(x) d\lambda(\alpha) \equiv \int_A h(\alpha) d\lambda(\alpha),$$

then (II-10) still holds. In particular, if $h(\alpha)$ and $\pi(\alpha,x)$ are such that $d\lambda(\alpha) = \delta'(\alpha - \alpha_o) d\alpha$[†] is admissible, then with α_o taken as the true value of α we obtain the C-R bound for unbiased estimates of $g(\alpha)$.

[†]The notation $\delta'(.)$ denotes the derivative of the Dirac distribution. See Reference [37].

Secondly, if the parameter space is not composed of a countable set of points one could perform a dissection of the parameter space into say, n points and cast the problem into the discrete form. Then one may solve for the suitable constants $\{a_i\}$ in (II-22). By taking n large enough it may be possible to obtain a good approximation to the results for the continuous parameter problem. The Barankin estimator which results from this procedure will be unbiased for every point in the dissection if a solution to (II-22) exists.

In arriving at the Barankin estimator for $g(\alpha)$ it has been noted that in general $T_o(x)$ will depend explicitly on α_o. Such an estimator is termed "locally best" at $\alpha = \alpha_o$. If only a locally best unbiased estimator for $g(\alpha)$ exists, it is not legitimate to choose α_o to correspond to the true value of α for the purpose of constructing the Barankin estimator, because then the Barankin estimator will explicitly depend on the parameter to be estimated. In this case, one must accept something less than an estimate which is best in the mean square sense about the true parameter value. In lieu of choosing a different approach entirely, it might be satisfactory to use the Barankin theory to construct an estimate which is best in the mean square sense about some arbitrarily chosen point α_o, which is not necessarily the true value of α. (Implicitly we are saying that the selection of α_o is made

from data which is statistically independent of the data to be used in constructing the Barankin estimator.) In any case, we will obtain an unbiased estimate of $g(\alpha)$ in this fashion if one exists.

We might then hope that the Barankin estimator, T_o, which is best about α_o does not differ widely from the estimator, T_1, which would have been obtained had α_o been chosen as the true value of α. As a measure of the sensitivity of T_o to the choice of α_o we could use

$$S = \frac{E\{T_o - T_1\}^2}{(\alpha_o - \alpha_1)^2} \tag{II-37}$$

where the expectation is taken over the distribution corresponding to α_1, the true value of α.

Then,

$$S \leq \frac{E\{\phi_o - \phi_1\}^2}{(\alpha_o - \alpha_1)^2} + \frac{\left[g(\alpha_o) - g(\alpha_1)\right]^2}{(\alpha_o - \alpha_1)^2} \tag{II-38}$$

In principle, once $d\lambda_o^o$ and $d\lambda_1^o$, (or their counterparts for the discrete parameter space) are found satisfying

$$\int_{A_{\alpha_o}} G(\alpha,\alpha'|\alpha_o)d\lambda_o^o(\alpha') = g(\alpha) - g(\alpha_o) \; ; \; \forall \alpha \epsilon A \tag{II-39}$$

$$\int_{A_{\alpha_1}} G(\alpha,\alpha'|\alpha_1)d\lambda_1^o(\alpha') = g(\alpha) - g(\alpha_1) \; ; \; \forall \alpha \epsilon A \tag{II-40}$$

a bound for S can be calculated.

This concludes the present discussion of the Barankin theory. We remark that the extension of the above to the treatment of vector parameter estimation problems is immediate upon replacing α with $\underline{\alpha} = [\alpha_1, \ldots, \alpha_n]$ and applying the above notions to each component of $\underline{\alpha}$.

APPENDIX III

In this appendix we present the actual computer programs used in the simulation of Examples 1 and 2.

The first listing pertains to Example 1, Case 1. To obtain the other cases for Example 1, the values of B0 and W0 (which correspond to β_o and ω_o) were changed.

The second listing pertains to the simulation of Example 2, Case 1, where only the center-frequency covariance parameter was treated as unknown.

As indicated on the note for the third listing presented, these statements appropriately modify the program for Example 2, Case 1 in order to simulate Example 2, Case 2.

```
//MAIN44     EXEC  FORTRAN
C     SIMULATION USING LEAST-SQUARES ESTIMATE
C
      IMPLICIT REAL*4 (M)
C     N1=NO. OF MONTE-CARLO TRIALS
C     N2=NO. OF TIME SAMPLES
C     T=OBSERVATION LENGTH
C     A0=UNKNOWN REGRESSION PARAMETER
C     B0=UNKNOWN COVARIANCE PARAMETER
C     G1=VECTOR OF KNOWN REGRESSION FUNCTION VALUES
C     R=DATA VECTOR
C     U=VECTOR OF INDEP. GAUSS. R.V'S
C     E=VECTOR OF CORRELATED GAUSS. R.V.'S
C
      WRITE(6,10)
   10 FORMAT(1H1,' SIMULATION FOR L.S. ESTIMATOR OF ALPHA'//)
C
C     THE PROGRAM WILL BE LIMITED TO A MAX OF TEN TIME SAMPLES
C
C
      COMMON KRANDG
      DIMENSION U(10),E(10),G1(10),R(10),PHI0(10,10)
      DIMENSION RHOHAT(1), AHAT1(1), AHAT2(1)
      DOUBLE PRECISION T,B0,RHO,B
C
C     INPUT INFORMATION
      N1=1000
      N2=10
      T=1.0
      B0=.01
      A0=2.0
      KRANDG=1
      IRAND=111111111
C
C     SOME AUXILLIARY CONSTANTS
      PI=3.1415926
      W0=100*PI
      F0=W0/(2*PI)
      DELTA=T/(N2-1)
C
      WRITE(6,15)N1,N2,T,B0,A0,F0
   15 FORMAT(1H0,' N1=',I6,'   N2=',I4,'   T=',F5.1,'  B0=',F4.2,
     1'  A0=',F4.2,'  F0=',F8.5,' CPS'//)
C
C     FORMING THE VECTOR G1
      DO 40 K=1,N2
      L=K-1
   40 G1(K)=COS(W0*DELTA*L)
C
C     PRE-COMPUTATIONS FOR L.S. ESTIMATE OF A0
      Y1=0.0
      DO 55 I=1,N2
   55 Y1=G1(I)**2+Y1
      T1=Y1
C
C     THE VARIANCE OF AHAT1 MAY BE CALCULATED ANALYTICALLY AND
C     WILL BE COMPARED WITH THE SIMULATION RESULTS
C
C     FIRST WE CALCULATE THE MINIMUM VARIANCE FOR
C     ESTIMATION OF A0--VAMIN
```

```
C
      RHO=DEXP(-BC*DELTA)
      Z=(G1(1)**2)*(1.-RHO**2)
      DO 60 I=2,N2
   60 Z=(G1(I)-RHO*G1(I-1))**2+Z
      VARMIN=(1.-RHO**2)/Z
C
C     SETTING UP THE TRUE COV. MATRIX,PHIO
      DO 100 I=1,N2
      DO 100 J=1,N2
      PHIO(I,J)=RHO**(J-I)
      PHIO(J,I)=PHIO(I,J)
  100 CONTINUE
C
C     CALCULATING THE VARIANCE OF THE L.S. ESTIMATE
C
      W=0.0
      DO 90 I=1,N2
      DO 90 J=1,N2
      W=G1(I)*PHIO(I,J)*G1(J)+W
   90 CONTINUE
      VLS=W/(T1**2)
C
C     MONTE-CARLO SIMULATION
C
      B=DSQRT(1.-RHO**2)
      X1=0.0
      X2=0.0
      X3=0.0
      Y7=0.0
      Y8=0.0
      Y9=0.0
      Y10=0.0
      Y11=0.0
      N=1
      DO 500 N5=1,N1
C
C     GENERATING INDEP. GAUSS. R.V.'S
      DO 20 L=1,10
      U(L)=RANUG(IRAND)
   20 CONTINUE
C
C     GENERATING THE DESIRED SAMPLED RANDOM PROCESS
      E(1)=U(1)
      DO 30 K=2,10
      E(K)=B*U(K)+RHO*E(K-1)
   30 CONTINUE
C
C     GENERATING THE DATA VECTOR
      DO 50 I=1,N2
      R(I)=A0*G1(I)+E(I)
   50 CONTINUE
C
C     THE L.S. ESTIMATE---AHAT1
      Y2=0.0
      DO 65 I=1,N2
      Y2=G1(I)*R(I)+Y2
   65 CONTINUE
      AHAT1(N)=Y2/T1
C
```

```
C       ESTIMATION OF RHO
        Y3=0.0
        DO 70 I=2,N2
        Y3=(R(I)-AHAT1(N)*G1(I))*(R(I-1)-AHAT1(N)*G1(I-1))+Y3
   70 CONTINUE
        Y4=0.0
        IN=N2-1
        DO 75 I=1,IN
        Y4=(R(I)-AHAT1(N)*G1(I))**2+Y4
   75 CONTINUE
        RHOHAT(N)=Y3/Y4
C
C       THE REVISED ESTIMATE---AHAT2
        Y5=R(1)*G1(1)*(1.-RHOHAT(N)**2)
        DO 200 I=2,N2
  200 Y5=(R(I)-RHOHAT(N)*R(I-1))*(G1(I)-RHOHAT(N)*G1(I-1))+Y5
C
        Y6=(G1(1)**2)*(1.-RHOHAT(N)**2)
        DO 210 I=2,N2
  210 Y6=(G1(I)-RHOHAT(N)*G1(I-1))**2+Y6
        AHAT2(N)=Y5/Y6
C
C       RE-ESTIMATION OF RHO
        X4=0.0
        DO 220 I=2,N2
        X4=(R(I)-AHAT2(N)*G1(I))*(R(I-1)-AHAT2(N)*G1(I-1))+X4
  220 CONTINUE
        X5=0.0
        DO 230 I=1,IN
        X5=(R(I)-AHAT2(N)*G1(I))**2+X5
  230 CONTINUE
        RHOHAT(N)=X4/X5
C
C       THE ITERATED ESTIMATOR OF ALPHA---AHATIT
        X6=R(1)*G1(1)*(1.-RHOHAT(N)**2)
        DO 240 I=2,N2
  240 X6=(R(I)-RHOHAT(N)*R(I-1))*(G1(I)-RHOHAT(N)*G1(I-1))+X6
        X7=(G1(1)**2)*(1.-RHOHAT(N)**2)
        DO 260 I=2,N2
  260 X7=(G1(I)-RHOHAT(N)*G1(I-1))**2+X7
        AHATIT=X6/X7
C
C       SAMPLE MEANS
        Y7=Y7+RHOHAT(N)
        Y8=Y8+AHAT1(N)
        Y9=Y9+AHAT2(N)
C       MEAN-SQUARED ERROR
        X1=(AHAT1(N)-A0)**2+X1
        X2=(AHAT2(N)-A0)**2+X2
        X3=(RHOHAT(N)-RHO)**2+X3
C       MEAN AND MSE OF ITERATED ESTIMATE
        Y10=Y10+AHATIT
        Y11=Y11+(AHATIT-A0)**2
  300 CONTINUE
        RHO=Y7/N1
        A1=Y8/N1
        A2=Y9/N1
C
        MSEA1=X1/N1
        MSEA2=X2/N1
```

```
      MSEKHO=X3/N1
C
      MAIT=Y10/N1
      MSEAIT=Y11/N1
      PERF=VAMIN/MSEAIT
C
      WRITE(6,7)
    7 FORMAT(1H0,'  A0     A1      A2      VLS      MSEA1     MSEA2'
     1'     VAMIN      MKHO     MSERHO      MAIT  MSEAIT      PERF',//)
C
      WRITE(6,12)A0,A1,A2,VLS,MSEA1,MSEA2,VAMIN,MKHO,MSEKHO,
     1MAIT,MSEAIT,PERF
   12 FORMAT(1H ,F4.1,2X,2(F6.4,2X),F6.3,2X,2(F7.4,2X),F7.4,3X,
     1F6.4,2X,F8.4,4X,F6.4,2X,F6.4,4X,F8.6)
      WRITE(6,14)IRAND
   14 FORMAT(1H0,' IRAND= ',I10)
      END
C     SUBROUTINE FOR GENERATING GAUSSIAN RANDOM VARIABLES
      FUNCTION RANDG(IRAND)
      GO TO (1,2), KRANDG
    1 U1=RANDM(IRAND)
      U2=RANDM(IRAND)
      XMAG=SQRT(-2.*ALOG(U1))
      X1=XMAG*COS(6.283184*U2)
      RANDG=X1
      KRANDG=2
      GO TO 20
    2 X2=XMAG*SIN(6.283184*U2)
      RANDG=X2
      KRANDG=1
   20 RETURN
      END
/*
/&       END OF JOB
```

```
//MAIN44     EXEC  FORTRANH,BCD
      IMPLICIT REAL*4 (M)
      CALL CLOCK(XKK)
      KKK=XKK
C     SIMULATING THE CASE FOR HETEROSKEDASTIC NOISE
C
C     N1=NO. OF MONTE-CARLO TRIALS
C     N2=NO. OF TIME SAMPLES
C     T=OBSERVATION LENGTH
C     A0=UNKNOWN REGRESSION PARAMETER
C     B0=NOISE BANDWIDTH PARAMETER
C     WC=NOISE CENTER FREQUENCY PARAMETER
C     W0=REGRESSION FUNCTION FREQUENCY (KNOWN)
C     G1=KNOWN REGRESSION VECTOR
C     R=DATA VECTOR
C     U=VECTOR OF INDEP. GAUSS. R.V.'S
C     V=AUXILIARY VECTOR OF INDEP. GAUSS. R.V.'S
C     E=VECTOR OF CORRELATED GAUSS. R.V.'S
C
C
      COMMON KRANDG
      DIMENSION U(10),E(10),G1(10),R(10),V(2)
      DIMENSION PHI0(10,10),PHIHAT(10,10),L1(10),L2(10)
      DOUBLE PRECISION PHI0,PHIHAT,B0,WC,AA0,BB0,CC,W0,W1,C3,C4,
     1C0,C1,C2,C5
      DOUBLE PRECISION RHO
C     INPUT INFORMATION
      N1=1000
      N2=10
      T=1.0
      B0=.01
      A0=2.0
      W0=0.0
      KRANDG=1
      IT=1
      IRAND=1111111111
      PI=3.1415926
      WC=3*PI
      FC=WC/(2*PI)
      F0=W0/(2*PI)
      DELTA=T/(N2-1)
C
      WRITE(6,15)N1,N2,T,B0,A0,F0,FC
   15 FORMAT(1H0,' N1=',I6,'   N2=',I4,'   T=',F5.1,'  B0=',F4.2,
     1'  A0=',F4.2,'  F0=',F6.3,' CPS    FC=',F6.3,' CPS'//)
C
C     FORMING THE VECTOR G1
      DO 40 K=1,N2
      L=K-1
   40 G1(K)=COS(W0*DELTA*L)
C
C     PRE-COMPUTATIONS FOR L.S. ESTIMATE
      Y1=0.0
      DO 55 I=1,N2
   55 Y1=G1(I)**2+Y1
      T1=Y1
C
C     SETTING UP THE TRUE COVARIANCE MATRIX
      DO 100 I=1,N2
      DO 100 J=1,N2
```

```
      PHIO(I,J)=DEXP(-BO*DELTA*(J-I))*DCOS(WC*DELTA*(J-I))
      PHIO(J,I)=PHIO(I,J)
  100 CONTINUE
C
C     CALCULATING THE VARIANCE OF THE L.S. ESTIMATE
      W=0.0
      DO 90 I=1,N2
      DO 90 J=1,N2
      W=G1(I)*PHIO(I,J)*G1(J)+W
   90 CONTINUE
      VLS=W/(T1**2)
C
C     CALCULATING REQUIRED CONSTANTS TO GENERATE THE DESIRED
C     TIME SERIES
C
      AAO=PHIO(1,2)*(1.-DEXP(-2*BO*DELTA))
      BBO=1.-DEXP(-4*BO*DELTA)
      CC=(BBO+DSQRT(BBO**2-4*AAO**2))/2.
      CO=DSQRT(1-CC**2)
      HO=DSQRT(CC)
      C3=HO*AAO/CC
      C4=DEXP(-2*HO*DELTA)
      C5=2*PHIO(1,2)
      H1=HO*C5-C3
      C1=(PHIO(1,2)-HO*H1)/CO
      C2=1.-(HO**2+H1**2)-C1**2
      C2=DSQRT(C2)
C
C     FOR THE REST OF THE PROGRAM WE NEED THE MATRIX INVERSE
C     OF PHIO--NOT PHIO.
C
C     THE FOLLOWING MATRIX INVERSION REPLACES PHIO WITH ITS INVERSE.
C
      CALL DPINV(PHIO,N2,TEST,DETER,L1,L2,N2)
C
C     WE NOW CALCULATE THE MINIMUM POSSIBLE VARIANCE
C     FOR ESTIMATING A0----VAMIN
C
      W1=0.0
      DO 95 I=1,N2
      DO 95 J=1,N2
      W1=G1(I)*PHIO(I,J)*G1(J)+W1
   95 CONTINUE
      VAMIN=1./W1
C
C
C     MONTE CARLO SIMULATION
C
      X1=0.0
      X2=0.0
      X3=0.0
      Y7=0.0
      Y8=0.0
      Y9=0.0
C
      DO 500 N5=1,N1
C
C     GENERATING THE SETS OF INDEP. GAUSS. R.V.'S
      DO 20 L=1,10
      U(L)=RANDG(IRAND)
```

```
   20 CONTINUE
      DO 21 L=1,2
      V(L)=RANDG(IRAND)
   21 CONTINUE
C
C     GENERATING THE DESIRED SAMPLED R.P.
      E(1)=H0*U(1)+C0*V(1)
      E(2)=H0*U(2)+H1*U(1)+C1*V(1)+C2*V(2)
      DO 30 K=3,N2
      E(K)=C5*E(K-1)-C4*E(K-2)+H0*U(K)-C3*U(K-1)
   30 CONTINUE
C
C     GENERATING THE DATA VECTOR
      DO 50 I=1,N2
      R(I)=A0*G1(I)+E(I)
   50 CONTINUE
C
C     THE L.S. ESTIMATE-----AHAT1
      Y2=0.0
      DO 65 I=1,N2
      Y2=G1(I)*R(I)+Y2
   65 CONTINUE
      AHAT1=Y2/Y1
C
C     ESTIMATION OF THE COVARIANCE MATRIX
C
C     THIS PROGRAM TREATS BANDWIDTH AS A KNOWN PARAMETER
C     WE ESTIMATE PHI0(1,2) AND THEN  THETA=WC*DELTA
      AHAT=AHAT1
   73 Y3=0.0
      DO 70 I=2,N2
      Y3=(R(I)-AHAT*G1(I))*(R(I-1)-AHAT*G1(I-1))+Y3
   70 CONTINUE
      Y4=0.0
      IM=N2-1
      DO 75 I=1,IM
      Y4=(R(I)-AHAT*G1(I))**2+Y4
   75 CONTINUE
      PHIHAT(1,2)=Y3/Y4
      THETA=ARCOS(SNGL(PHIHAT(1,2))*SNGL(DEXP(B0*DELTA)))
C
C     SETTING UP THE ESTIMATED COVARIANCE MATRIX
      DO 300 I=1,N2
      DO 300 J=I,N2
      PHIHAT(I,J)=DEXP(-B0*DELTA*(J-I))*COS(THETA*(J-I))
      PHIHAT(J,I)=PHIHAT(I,J)
  300 CONTINUE
C
C     THE REVISED ESTIMATE-----AHAT2 OR AHAT1T
      CALL DPINV(PHIHAT,N2,TEST,DETER,L1,L2,N2)
      W2=0.0
      DO 310 I=1,N2
      DO 310 J=1,N2
      W2=G1(I)*PHIHAT(I,J)*G1(J)+W2
  310 CONTINUE
      W3=0.0
      DO 315 I=1,N2
      DO 315 J=1,N2
      W3=G1(I)*PHIHAT(I,J)*R(J)+W3
  315 CONTINUE
```

```
      AHAT=W3/W2
      IF(IT-1) 63,63,64
   63 AHAT2=AHAT
      IT=2
      GO TO 73
   64 AHATIT=AHAT
      IT=1
C
C     PERFORMANCE CALCULATIONS
C
C     SAMPLE MEANS
      Y7=Y7+AHAT1
      Y8=Y8+AHAT2
      Y9=Y9+AHATIT
C
C     MEAN-SQUARED ERROR
      X1=(AHAT1-A0)**2+X1
      X2=(AHAT2-A0)**2+X2
      X3=(AHATIT-A0)**2+X3
      CALL CLOCK(KLL)
      KKM=KLL
      IF(IABS(KKM-KKN).GE. 25200) GO TO 1500
  500 CONTINUE
 1500 CONTINUE
      MA1=Y7/N5
      MA2=Y8/N5
      MAIT=Y9/N5
      MSEA1=X1/N5
      MSEA2=X2/N5
      MSEAIT=X3/N5
      PERF=VAMIN/MSEAIT
      WRITE(6,7)
    7 FORMAT(1HO,'  A0     MA1      MA2      VLS      MSEA1    MSEA2'
     1'     VAMIN      MAIT     MSEAIT      PERF',//)
C
      WRITE(6,12)A0,MA1,MA2,VLS,MSEA1,MSEA2,VAMIN,MAIT,MSEAIT,
     1PERF
   12 FORMAT(1H ,F4.1,2X,2(F6.4,2X),F6.3,2X,2(F7.4,2X),F7.4,4X,
     1F6.4,2X,F6.4,4X,F8.6)
      WRITE(6,1100) N5
 1100 FORMAT(1H ,' ACTUAL NO. OF ITERATIONS= ',I8/)
      WRITE(6,14) IRAND
   14 FORMAT(1HO,' IRAND= ',I10)
      END
/*
/&            END OF JOB
```

```
//MAIN44     EXEC  FORTRAN
C      HERE BANDWIDTH AND CENTER-FREQUENCY ARE REGARDED AS UNKNOWN
C      WE ESTIMATE  PHIO(1,2) AND  PHIO(1,3) AND THEN
C      THETA=WC*DELTA  AND  EXP(-BO*DELTA)=RHO
       AHAT=AHAT1
   73  Y3=0.0
       Y4=0.0
       Y5=0.0
       DO 70 I=2,N2
       Y3=(R(I)-AHAT*G1(I))*(R(I-1)-AHAT*G1(I-1))+Y3
       Y5=(R(I)-AHAT*G1(I))**2+Y5
   70  CONTINUE
       DO 71 I=3,N2
       Y4=(R(I)-AHAT*G1(I))*(R(I-2)-AHAT*G1(I-2))+Y4
   71  CONTINUE
       PHIHAT(1,2)=Y3/Y5
       PHIHAT(1,3)=Y4/(Y5-(R(2)-AHAT*G1(2))**2)
       THETA=ATAN(SNGL(DSQRT(DABS(1.-PHIHAT(1,3)/PHIHAT(1,2)**2))))
       RHO=PHIHAT(1,2)/COS(THETA)
C
C      SETTING UP THE ESTIMATED COVARIANCE MATRIX
       DO 300 I=1,N2
       DO 300 J=1,N2
       PHIHAT(I,J)=(RHO**(J-I))*COS(THETA*(J-I))
       PHIHAT(J,I)=PHIHAT(I,J)
  300  CONTINUE
/*
/&        END OF JOB
```

Note: These statements replace statements 73 through 300 in the previous listing when it is desired to simulate Case 2 of Example 2.

Security Classification

DOCUMENT CONTROL DATA – R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
University of Southern California
Department of Electrical Engineering
Los Angeles, California 90007

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE
ESTIMATING LINEAR REGRESSION PARAMETERS IN THE PRESENCE OF NON-WHITE GAUSSIAN NOISE WITH UNKNOWN COVARIANCE PARAMETERS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Scientific Interim

5. AUTHOR(S) *(First name, middle initial, last name)*
Gene Barry Goldstein

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1968	114	55

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF-AFOSR 68-1555	
b. PROJECT NO. 9749-01	
c. 6144501F	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* AFOSR 69-0319TR
d. 681304	

10. DISTRIBUTION STATEMENT
1. This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH, OTHER	Air Force Office of Scientific Research (SRMA) 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

A number of problems that arise in radar and sonar applications can be regarded as parameter estimation problems, in which the desired signal, $f(t, \underline{\alpha})$, is imbedded in non-white, Gaussian noise. It is desired to estimate the unknown, nonrandom parameter vector, $\underline{\alpha}$, from observations (continuous or sampled) of the received noisy signal over a finite time interval $[0, T]$. Here $f(t, \underline{\alpha})$ is a known nonstochastic function, and we shall consider the case when $f(t, \underline{\alpha})$ is linear in $\underline{\alpha}$. In this case, $\underline{\alpha}$ is referred to as a linear regression vector "(U)"

We shall investigate the variance of the Least-Square (LS) estimator and of the so-called Generalized-Least-Squares (GLS) estimator for $\underline{\alpha}$. Both are unbiased estimators for $\underline{\alpha}$. "(U)"

When the noise covariance function is completely known one may construct a minimum variance unbiased estimator (MVUE) for $\underline{\alpha}$, and this estimator is a member of the class of GLS estimators. "(U)"

Our interest is in the case when the noise covariance is not completely known, but may be regarded as a known function of a finite number of unknown, nonrandom parameters, $\underline{\beta}$. "(U)"

It is shown that when $\underline{\beta}$ contains any covariance parameters other than the noise variance, there exists no MVUE for $\underline{\alpha}$. "(U")

However, we shall exhibit a class of problems for which the MVUE for $\underline{\alpha}$ has a variance which is orders of magnitude smaller than that of the LS estimator. In such a case it is of interest to find an estimator which makes use of whatever covariance information is available in an attempt to approach the performance of the MVUE. "(U)"

DD FORM 1 NOV 68 1473

Security Classification

Item 13 Abstract (con't)

It is shown that we can significantly improve upon the LS estimator by employing a bootstrapping procedure to estimate $\underline{\alpha}$. In some cases the bootstrapped estimate of $\underline{\alpha}$ can be shown to be unbiased. In any case, it is demonstrated via computer simulation that the bootstrapped estimate of $\underline{\alpha}$ is capable of reducing the variance of the LS estimate by orders of magnitude. In fact, the mean squared estimation error using the bootstrapped estimator for $\underline{\alpha}$ may be within a few percent of the variance of the MVUE, i.e., the variance the MVUE would have if $\underline{\beta}$ were known a-priori. "(U)"

The bootstrapping procedure consists of using the LS estimate of $\underline{\alpha}$ to provide an initial estimate of the regression vector from which an initial estimate of the unknown covariance parameters is constructed. "(U)"

Two procedures are outlined to accomplish the estimation of $\underline{\beta}$. The first approach is based upon an application of the theory of locally best unbiased estimation. The second approach is herein termed the "inverse-covariance-function" technique. Because of its simplicity, the latter approach is employed in the simulations. "(U)"

Regardless of the manner in which the covariance parameters are estimated, these estimates are used to construct the GLS estimator for $\underline{\alpha}$. This is the first iteration of the bootstrapping procedure. "(U)"

The GLS estimate of $\underline{\alpha}$ is then used to re-estimate the unknown covariance parameters, and then to re-estimate the regression parameters. "(U)"

The process uses only the one available record of data, and may be repeated ad nauseam. However, dramatic results were obtained after only two iterations of the bootstrapping procedure. "(U)"

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Estimating linear regression parameters Non-white Gaussian noise Unknown covariance parameters Nonrandom parameter vector, α						